

A culture whose time is past

France's scientific and technological structures have no choice but to undergo a revolution. The government must encourage a climate of competitiveness and risk-taking required to ensure a strong scientific future.

Forgetting the economic disasters that were Concorde and Superphénix, even the most ardent free marketeer must concede that France's model of industrial policy, anchored in a system of meticulous and comprehensive state planning, has borne fruits. But the model has now outlived much of its usefulness. France's *grands programmes technologiques* were well suited to developing space, aeronautics and nuclear power. But the technologies that will drive tomorrow's economy, such as biotechnology and informatics, demand imaginative entrepreneurship, a tough competitive streak and, above all, the flexibility and speed needed to bring the latest research to the marketplace.

This message has once again been reinforced in a damning report to the government (see page 214), which states bluntly that France's industrial and research systems are poorly prepared to take up this challenge, despite a solid basic science base. If the government of Lionel Jospin is to fulfil its pledge to make a technology and innovation policy the centrepiece of its efforts to make France competitive in the twenty-first century, it will have its work cut out. The challenge facing this alliance of socialists, communists and greens is to transform France's industrial culture without the social disruption characteristic of a Thatcherite revolution. Whether it has the stomach for the task is a critical question for the country.

Attacking the many administrative and other obstacles that dog innovation could certainly bring some improvements. But substantial change will require nothing short of abandoning France's traditional technocratic and hierarchical approach to innovation in favour of a 'bottom-up' strategy aimed at empowering the small and medium-sized companies that all agree are the key to tomorrow's economic successes. These have been almost neglected so far, with almost all state industrial research monies, for example, going to a handful of large companies associated with the *grands programmes*.

With defence spending declining, there is now scope for transferring much more state support to small companies in promising sec-

tors. The report's recommendation that an interministerial body be set up to monitor and evaluate a strategy for deciding where public funds for stimulating industrial research are best spent makes good sense. That such a body does not already exist is itself astonishing. As the report aptly questions, might not the FF820 million of public money used to create BioAvenir, a joint research programme between Rhône Poulenc and research agencies, have been better spent, for example, on encouraging the creation of biotechnology start-ups?

The state certainly has a role to play, but only if it provides scientists with the opportunity to explore the commercial viability of new ideas free of the straitjackets of bureaucracy on the one hand and short-term profitability on the other. The last thing the entrepreneurs behind Génopole, France's planned biotechnology valley, want is to be sitting in meetings with ministry bureaucrats deciding how subsidies should be spent while their competitors push ahead.

The report's recommended creation of a halfway house between research laboratories and industry, along the lines of Germany's Fraunhofer Institutes, is a sensible task for the state, and would fill an obvious gap in the country's technology infrastructure. But what is ultimately needed is less state involvement, not more. General Charles De Gaulle is no doubt turning in his grave at the many references in the report holding up the United States and United Kingdom as models of how to encourage innovation. But it is hard to avoid the conclusion that more competitiveness and less job security are necessary, in both research and industrial bases.

Next month, Lionel Jospin will chair an interministerial council aimed at drawing up a strategy and priorities for French research. A firm commitment to eliminating restrictive state interference and bureaucracy wherever possible, and to introducing concrete measures aimed at creating a more competitive but also more productive environment for high-technology companies and their investors, is necessary for France's economic health. The long-term health of France's basic research also depends on it. □

Competitiveness versus conferences?

A small rebellion reflects postdocs' dissatisfaction with meetings of questionable value.

Once upon a time there was a keen young researcher working in a rapidly developing field. After months of hard work he at last had something interesting to tell his peers. He registered for what looked like an excellent conference at which to present his results: international, supported by an international agency, many of the top people attending. Somewhat taken aback when he discovered the costs, nevertheless, eventually, he went. The accommodation was at a comfortable hotel, the meeting facilities excellent. He was appalled. The talks were mainly reviews or results whose publication was imminent anyway. He discovered that the field was too competitive to allow people to discuss their latest findings. All that money, and for what?

That was the experience of at least one postdoc who attended a European Research Conference on programmed cell death (apoptosis) last year. The costs of attendance at the three-day event, not

including travel, amounted to DM1,800 (about \$900), despite financial support for the conference by the European Union.

Given what they perceive to be the poor value of what was delivered, and a more general trend in this direction, a group of postdoctoral researchers has now rebelled and organized a very different meeting at cheap rates in an Italian research institute. Thanks to sponsorship, the fee to those invited will be \$25 — and participants will have to pay only travel expenses.

The topic will again be apoptosis, and the organizers promise to set an example by talking about their latest results (e-mail contact: delaurenz@utovrm.it). These are serious scientists, and the meeting promises to be stimulating.

This initiative represents a statement by a few postdoctoral researchers about the state of conferences in at least one fast-moving area of science. Is it the tip of an iceberg of dissatisfaction? □



German Greens soften views on biotech as power edges closer

[MUNICH] Sensing a possible electoral victory in Germany's federal elections in September, which could result in a coalition government with the Social Democrat Party (SPD), the Greens have relaxed their critical approach to biotechnology and genetic engineering.

At the same time, German scientists are reacting cautiously to reports that, if such a coalition were to come to power, the Greens might choose the science ministry as one of the three ministries they would run (the other two being the foreign and the environment ministries).

At the Greens' annual conference, held in Magdeburg earlier this month, a large majority of delegates voted in favour of a motion supporting genetic engineering in drug design and diagnostics, as well as basic research. They agreed to add this so-called *Gentechnikbeschluss* to the party's election programme, but remain opposed to the use of genetic engineering in agriculture.

According to the news magazine *der Spiegel*, Gerhard Schröder, the SPD candidate for chancellor, is already considering Krista Sager as possible research minister in a coalition cabinet. Sager, who is 44, is science minister in the *Land* (state) of Hamburg. She is both the first woman and the first Green to head a science ministry in Germany.

Many scientists and industry lobbyists are worried about such a prospect, given the Greens' fundamentalist grassroots. "We are not panicking, but if the Greens expect us to give up what we have achieved, for example in the Human Genome Project, they will face bitter resistance from scientists," says Detlev Ganten, head of the Max-Delbrück Centre for Molecular Medicine in Berlin.

Nevertheless Ganten, who is also president of the Helmholtz Society, the umbrella organization of Germany's national research centres, admits the Greens have abandoned many of their extreme positions, and believes they will reject more if they take on government responsibility. He describes Sager as "pragmatic, open-minded and competent".

Furthermore, politicians from either side insist that no definite decisions about ministerial positions will be made before the election. If an SPD-Green coalition is formed, Sager will have to compete with Social Democrat rivals.

In its effort to project a modernized image, the SPD has enthusiastically taken up the cause of research and technology, and in its election programme promises to double the current DM15 billion (US\$8 billion) annual federal budget for education,



Sager: could be next research minister.

research and technology. For their part the Greens, criticizing the promise as unrealistic, have proposed a comparatively modest increase of DM2 billion.

Wolf-Michael Catenhusen, the SPD spokesman on research, and another possible candidate as research minister, says that unresolved conflicts within the Greens about topics including the whole range of gene technology, disqualify them from taking responsibility for research and technology policy. Despite the *Gentechnikbeschluss*, some Greens still want all biotechnological products and methods banned.

But Manuel Kiper, spokesman for the Greens' parliamentary group on research policy, argues that his party's ideas on research — such as its 'soft' approach to biotechnology — are in line with German public opinion.

In addition to their traditional demand to abandon nuclear energy, the Greens want to end Germany's contribution to various

major international research projects, including the proposed fusion reactor ITER and the international space station (and manned space exploration in general).

Kiper stresses, however, that his party accepts the general importance of science and technology, and says the Greens should no longer be described as a 'Stone Age party'. "We certainly don't continue to cling to Green positions that were formed 12 years ago," he says.

Under the framework of sustainable development and social responsibility, the main pillars of the Greens' science policy are environmental research, information technology and research on transport.

In contrast to the views of the present research minister, Jürgen Rüttgers, and to mainstream political thinking, the Greens say basic research should be given higher priority than applied research. This surprising stance first became apparent in the party's position paper on science policy, published more than a year ago (see *Nature* 385, 382; 1997). The paper criticized Rüttgers for "overemphasizing industrial research aimed at short-term economic gain" and called for the re-establishment of basic research as the main priority.

Quirin Schiermeier

Clinton told to speed up plutonium disposal

[WASHINGTON] A high-powered Washington panel has sharply rebuked the Clinton administration for dragging its feet on the disposal of plutonium from surplus US nuclear warheads.

The Center for Strategic and International Studies says that the administration has failed to get the disposal programme under way at home, and to win agreement for similar efforts in Russia.

A report from the centre calls for a formal agreement between the United States and Russia to ensure that plutonium 'pits' — or cores — from thousands of nuclear warheads rendered surplus by arms-control agreements are processed into safer shapes, and then disposed of.

But the panel of senior scientists, engineers and former diplomats who drew up the report are especially angry with the US administration for failing to make disposal a higher priority. "At the moment, there is more leadership in Russia on this than there is in the United States," says John Taylor, a consultant at the Electric Power Research Institute in California, and chairman of the panel.

Senator Pete Domenici (Republican, New

Mexico), who chairs the committee allocating the disposal programme's budget at the Department of Energy, co-chaired the panel and endorses its conclusions.

"This issue has not received the appropriate level of interest in the Clinton administration," says Domenici, who warns that Russian weapons components could fall into the hands of terrorists if inaction continues.

The report says that both the United States and Russia should press ahead immediately with converting the pits into less dangerously shaped blocks of plutonium, in advance of disposing of the plutonium itself.

The United States is building a prototype facility to do this at the Los Alamos National Laboratory in New Mexico, using a dry process based on hydrogen gas. But a full-scale facility is not due to come on line until 2005.

The report attributes the overall slow progress to a lack of commitment from senior administration officials. "There presently is little or no serious leadership within the US government... to move the programme forward," it says. Colin Macilwain

Innovation 'stifled' in France, says report

[PARIS] A decade of efforts by successive French governments to encourage technology transfer and wealth creation have had little impact, according to a highly critical report commissioned by the French government and released last week. The report paints a picture of a country in which innovation and entrepreneurship are stifled by bureaucracy, and whose state industrial policies are out of touch with the needs of fast-moving sectors such as biotechnology and computing.

The 300-page report was commissioned from Henri Guillaume, honorary president of the national innovation agency Anvar, by Claude Allègre, the minister for national education, research and technology, and Dominique Strauss-Kahn, the industry and finance minister. It catalogues a series of weaknesses and obstacles to wealth creation across the research and industrial system.

The report says that virtually the same diagnosis was made in a 1985 report commissioned from the OECD by the then science minister, Hubert Curien. The declared commitment of successive governments to wealth creation has not been translated into action, asserts Guillaume, adding that the state "lacks a proper strategy" for encouraging industrial research and evaluating the success of state innovation programmes.

By any yardstick, France lags behind the United States and Japan in industrial research, particularly in computing and biotechnology, according to the report, which says research agencies and universities



Guillaume: critical of industrial culture.

are too isolated both from each other and from industry. Despite attempts to encourage researchers to work in industry, each year only 30 to 40 of the 25,000 researchers working for public research agencies transfer to industry. Observers say this situation is unlikely to change overnight. They point out that, despite claims to the contrary by the research agencies, the French research system is still geared almost entirely to rewarding academic success, and industrial prowess is largely ignored in the evaluation of scientists.

Excessive bureaucracy plagues efforts to create wealth, says Guillaume, pointing out that universities that negotiate industrial contracts must agree the number of posts in advance and have the terms agreed by its executive board. The haphazard proliferation of government schemes — there are over a hundred sources financing innovation — has also led to confusion, he writes.

But although his report is long on criticism, it has been widely interpreted as being short on solutions, and many are also sceptical as to whether its recommendations will be successfully translated into action. "There is no miracle solution," admits Guillaume.

Jean-Bernard Le Pecq, who stepped down as research director of Rhône-Poulenc Rorer this month to set up his own gene therapy

company, Epicells, agrees. "It is difficult to identify single factors that could lead to major changes," he says.

The report recommends that the government clarify and streamline its industrial policies by creating a national technological centre to coordinate and monitor industrial activities throughout the research agencies and universities. A rigorous follow-up of innovation policies is needed, it says.

The report also proposes the creation of professional technology transfer centres within all universities, and the setting up of national networks of laboratories in important areas, to provide a clear interface between industry and the research agencies and universities.

One priority identified by researchers, which is taken up in the report, is the need to divert government subsidies to providing money for scientists to set up their own companies. Whereas venture capital for expanding existing companies is relatively plentiful, start-up funds to turn research ideas into new companies remain in short supply.

The report recommends the creation of national funds to provide such seed money in the biotechnology and computing sectors. It also proposes that tax credits for research, currently restricted to companies that implant in specified development zones, should be extended to the whole country.

Many observers agree with the recommendation that the government alter the laws forbidding publicly funded researchers from holding shares or sitting on the boards of companies with which they have research links. A bill that would do just that is now being considered by the government.

The current restrictions are a "fundamental obstacle" to innovation, says Pierre Tambourin, former head of life sciences at the Centre National de la Recherche Scientifique (CNRS), who is responsible for developing the 'Genopole' biotechnology park to be built by the renowned Génethon genome laboratory near Paris. "Unless the government makes a special dispensation for civil servant researchers we will be able to write the same report in another ten years time."

Tambourin says there has been a large shift in the attitudes of researchers towards industry since the 1980s, as the aversion to working with industry has given way to enthusiasm. But he argues that the main barrier to technology transfer remains the lack of industrial competence and an entrepreneurial culture among French scientists.

Substantial change requires the emergence of a business culture in the scientific community, says Tambourin. "French scientists know virtually nothing about industry, and most young recruits don't even know what a patent is."

Declan Butler

Allègre seeks boost for young researchers

[PARIS] Claude Allègre, the French minister for national education, research and technology, has told research agencies to come up with proposals within a month for increasing the independence of young researchers. He argues that the current system, under which funds are distributed to laboratories rather than individual teams, favours established researchers.

Allègre announced his plans last week after the cabinet had approved a package of measures to support young scientists. The government has created more than 6,000 new research posts this year, and intends to establish rolling five-year plans to manage

recruitment and rejuvenate the research population. The average age of researchers is 47½ — ten years more than in the United Kingdom, for example.

Allègre argues that young scientists must be given greater responsibility. A recent survey by the French embassy in Washington, says a major reason that French postdoctoral scientists stay in the United States, rather than returning home, is the greater autonomy they enjoy there.

Allègre also announced plans to let young physicians work in clinical research within research agencies without first having to do a PhD as they have to now. To increase mobility among

young scientists, the Centre National de la Recherche Scientifique and other research agencies would create new posts to take in university researchers.

The government will this year support the creation of posts for 650 engineers and 350 scientists within small and medium-sized companies, he said, and make agreements with more than 100 companies for industrial placements for researchers during PhDs. He promised that a bill to lift the obstacles preventing publicly funded researchers from having commercial ties (see above) would also contain provisions geared to helping young scientists to create their own companies. **D.B.**

Asteroid watchers debate false alarm

[WASHINGTON] A worldwide media frenzy last week about reports of a possible asteroid collision with the Earth in 2028 has led to scientists taking a close look at their strategy for informing the public about such events.

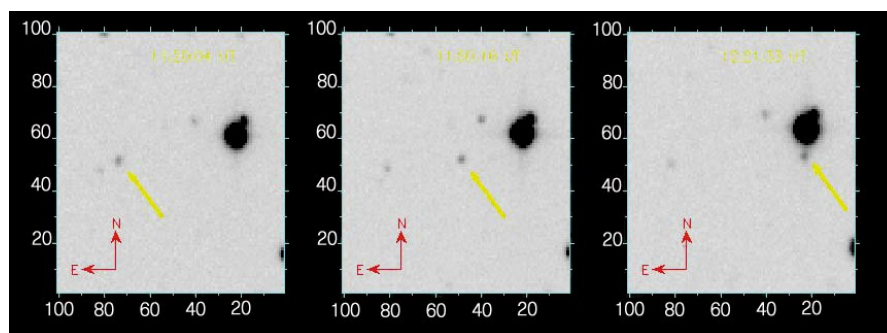
Many experts in the field are critical of a decision by the Central Bureau for Astronomical Telegrams in Cambridge, Massachusetts, to send out a press release warning of a possible collision before first checking with other researchers. But others say the false alarm was an inevitable by-product of the open way in which asteroid researchers share information.

The furore began on 11 March, when the bureau — a non-profit organization that serves as a clearing-house for astronomical alerts — e-mailed a 'circular' predicting that the mile-wide asteroid 1997 XF11 would pass very close to Earth in 2028, on the basis of observations made a week earlier. The circulars go out to subscribing astronomers, but some journalists also receive them.

Anticipating that the story would spread quickly, Stephen Maran, an astronomer at the US space agency NASA's Goddard Space Flight Center in Maryland, and long-time press officer for the American Astronomical Society (AAS), suggested to the bureau's director, Brian Marsden, that he send a more general release through the AAS to science reporters around the world.

Marsden, an expert on asteroid orbits, then issued a release including the statement: "The chance of an actual collision is small, but one is not entirely out of the question." The news of a potentially devastating asteroid impact made headlines around the world.

Ted Bowell of the Lowell Observatory in Arizona, whose search for near-Earth objects



Watch out: the first images of 1997 XF11 from Kitt Peak's Spacewatch telescope, 6 December 1997.

will soon join two other NASA-funded searches already under way, is among those who believe the warning was released prematurely. Marsden, he says, "should have consulted with several of his colleagues before rushing into print".

Eleanor Helin of NASA's Jet Propulsion Laboratory, who runs a search called NEAT, and Bowell turned up photographs of the asteroid from 1990 in less than a day, allowing more accurate orbit measurements. The new data placed the object a safe 600,000 miles from Earth at its closest approach.

"I don't think we were crying wolf," says Marsden, pointing out that the small community of asteroid researchers usually works this way, sharing orbital data and then asking colleagues to make new observations or search for archival data to improve the estimates. He also says that until Helin's photos surfaced, three of four groups calculating the asteroid's orbit agreed that there was a remote but 'non-zero' chance of a collision in 2028.

Helin agrees that the system of communication by circular generally works, and says "we couldn't have acted any faster".

David Morrison of the NASA Ames Research Center, who led a comprehensive assessment in 1992 of the asteroid threat, agrees "it would have been smarter for the community to seek those [other archived observations] before going public". But, he says, "I'd much rather that we had an open system, even if it might sometimes embarrass us, than that somebody try to control it."

Scientists involved in the Search for Extraterrestrial Intelligence (SETI) have addressed the problem by agreeing to a protocol whereby anyone who detects a suspected extraterrestrial signal would consult privately with other colleagues before going public with the news. This may not work, however, in the asteroid community, which has long-lasting rivalries. "Maybe the SETI people are more harmonious," jokes Marsden.

The flap about 1997 XF11 was expected to be a hot topic among asteroid researchers meeting this week in Houston, Texas, to discuss how the various campaigns to identify new asteroids are faring, and how the search might be accelerated.

The meeting, sponsored by NASA, had been planned before last week's events. Morrison's 1992 report called for spending approximately \$5 million a year for 10 years to survey the population of Earth-crossing objects more than one kilometre in diameter, 90 per cent of which have not yet been identified.

But NASA is currently spending only \$1 million a year. "At the current rate of discovery, it will take more than a century to find 90 per cent or more of the objects [the size of 1997 XF11] with Earth-crossing orbits," says Morrison.

Congress may call for stepping up the pace given the worldwide attention focused last week on the asteroid threat. The day after Marsden's announcement, the chairman of the House of Representatives space subcommittee, Dana Rohrabacher (Republican, California), called on President Bill Clinton to reverse his veto of the Pentagon-sponsored project Clementine 2, which would fire small interceptors into the surface of an asteroid (see *Nature* 389, 776; 1997). **Tony Reichhardt**

Decline in Indian publications raises alarm

[NEW DELHI] A steady decline in the international profile of Indian science has been revealed by an analysis of research papers published in recent years. This "is a cause of concern", says the study.

India has fallen from eighth largest publishing nation in 1989 to twelfth position, with just a 2 per cent share of the world's journal articles — a significant drop from the 2.8 per cent in 1989.

The study was carried out by Subbiah Arunachalam of the Indian Institute of Technology in Chennai and R. Srinivasan and Vidyalakshmi Raman of the Central Electrochemical Research Institute in Karaikudi, and the results are presented in the latest issue of *Current Science*, published by the Indian Academy of Sciences.

The authors analysed papers originating in India and published in 2,300 journals

indexed in the *Science Citation Index* (SCI) in 1989–92. Nearly a quarter of the papers from India are in chemistry, 18 per cent in physics, and 13 per cent in engineering. Biology accounted for 12 per cent.

Although India's share of the world's journal articles decreased marginally, there has been a significant increase in the number of papers published by Indian scientists in foreign journals as a whole, rising from 8,751 in 1989 to 9,958 in 1992. In contrast there was a 27 per cent decline in the number of papers published in the 12 Indian journals indexed in SCI.

India invested US\$1,460 million — 0.83 per cent of its gross national product — in research and development in science and technology in 1992–93, a figure that has been steadily declining from 0.93 per cent in 1989–90. **K. S. Jayaraman**

Canada's environment act under attack...

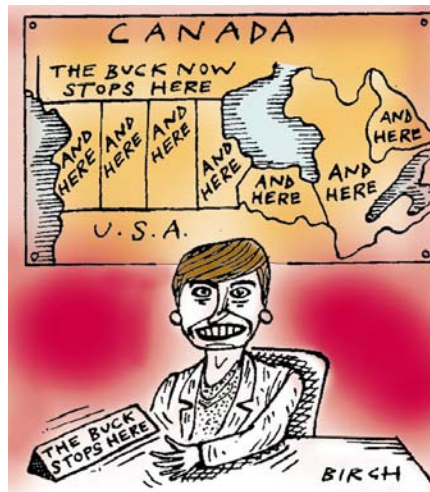
[MONTREAL] Canada's environment minister, Christine Stewart, last week tabled a revised Environmental Protection Act that she said would strengthen environmental protection in Canada — but that critics said would weaken it.

The legislation places emphasis on voluntary efforts by industry, and increased cooperation with the provinces. But environmentalists argue that such moves will undermine the federal environment department's traditional controls, and give the provinces a virtual power of veto over environmental protection measures.

One major innovation in the legislation is a National Accord on Environmental Harmonization, and three sub-agreements, each signed on 30 January by the environment minister and the provinces. These deal with environmental assessment, the establishment of national environmental standards and inspection activities under federal laws.

The federal agency Environment Canada hails the agreement as a significant advance. "Under the new legislation, the focus of environmental protection in Canada would shift from cleaning up after the damage is done to preventing pollution in the first place," Stewart said in a statement.

She says the legislation reflects some of the concerns that have recently been expressed, including the need for increased recognition of voluntary efforts by industry, and improved consultation with provinces and territories. "We have listened to all our partners and stakeholders, and this legislation represents a reasonable and balanced



approach to environmental protection in cooperation with all parties," says Stewart.

But critics are worried about central government ceding responsibility on environmental matters to the provinces. "The signing of the accord marks the end of any significant role for the federal government in the protection of the environment for the foreseeable future," says Shelly Bryant, executive director of Action: Environment.

Paul Muldoon, executive director of the Canadian Environmental Law Association, says the agreement "will make it virtually impossible to deal with the major environmental challenges facing Canada". He describes the implications for issues such as the implementation of the Kyoto protocol on global climate change as "simply horrendous". The association is asking the Federal

Court of Canada to overturn the agreement.

The act comes at a difficult time for both Environment Canada and the provincial environmental agencies. The federal body's budget has been cut by 34 per cent since 1994–95. The government's eventual goal is a 40 per cent reduction.

Besides its domestic pollution responsibilities, Environment Canada must participate in more than 20 international environmental agreements, including the Great Lakes Water Quality Agreement, the Basel Convention on Toxic Exports and the Agenda 21 pledge made at the United Nations Rio summit in 1992. Critics say its ability to do this has now been severely compromised.

Gary Gallon, president of the Canadian Institute for Business and Environment, which carries out economic analyses of the impact of environmental protection, says the cuts by federal and provincial governments are one reason the harmonization accord will not work.

In contrast to the Ontario environment ministry's official figures, he claims its operating budget has been reduced by 42 per cent between 1994–95 and 1997–98. He says Quebec's budget for environmental protection has dropped 64.9 per cent during the same period, Alberta's environment ministry budget has lost 31 per cent since 1992, and Newfoundland's has fallen 60 per cent since 1995.

Gallon claims the federal ministry and Ontario and Alberta provinces are committed to reducing their environmental regulations by up to 50 per cent. **David Spurgeon**

...as federal agency is accused of undermining pollution watchdog

[MONTREAL] Environment Canada is embroiled in a dispute that has seen critics accusing it of trying to interfere with the independence of the Commission for Environmental Cooperation (CEC), a body set up by Canada, the United States and Mexico under the North American Free Trade Agreement.

The commission had circulated a draft document listing prominent Canadian companies among North America's worst polluters. But François Laval, head of the pollutant release inventory of Canada's federal environment department, wrote to the CEC saying that Canada planned to take steps to ensure that the report was not released "unless Canadian concerns are addressed".

Laval warned that Christine Stewart, the environment minister, might refuse to endorse the report unless polluters were allowed first to review its findings. The dispute is said to be linked to the resignation last month of the CEC's executive director, Victor Lichtinger, a Mexican diplomat. He had just signed a three-

year contract, but left after firing the commission's senior US official, Greg Block.

Block was alleged to have leaked privileged information involving Canada to his country. The CEC said Lichtinger left voluntarily, but some critics claim that he was forced out because he had become too sympathetic to environmentalists' concerns.

"The Canadian government seems more interested in defending the image of polluters than in telling Canadians what they have a right to know," says Matthew Bramley of Greenpeace in Montreal. And Stewart Elgie, a lawyer who chairs a national committee of environmentalists and industry groups, warns of a "real risk" that the CEC could lose its political independence.

The government denies any political motivation, arguing that its main complaint was about the accuracy of the report and saying that the commission has agreed voluntarily to make amendments. In his letter to the CEC, Laval claimed the report was

misleading, unscientific and riddled with errors. But it was later learned that industry had provided erroneous figures to Environment Canada, which the agency simply passed on to the commission.

The contentious report, *Taking Stock*, is the second CEC document that has raised hackles. Last year the CEC identified Ontario, Canada's richest province, as the third-worst polluting jurisdiction in North America, behind Texas and Tennessee. Ontario's environment minister, Norman Sterling, rejected the commission's conclusions. He said Ontario ranks poorly because it is a large province with a large industrial base, and because its industries are more honest than those elsewhere in reporting discharges.

But an assessment by his own ministry, obtained by the Canadian Institute for Environmental Law and Policy under a Freedom of Information Act request, failed to uncover any errors in the report that would suggest Ontario was treated unfairly. **D.S.**

UK physics 'must focus on breakthroughs'

[LONDON] Physics and astronomy research in Britain should concentrate its resources on areas with the potential to create significant scientific breakthroughs, rather than distributing them more widely, suggests the incoming chief executive of Britain's Particle Physics and Astronomy Research Council (PPARC).

Ian Halliday, professor of physics at the University of Wales, Swansea, argues that years of budget cuts mean that scientists tend to opt for the security of well-trodden ideas. He wants this to change, and is willing to fund riskier projects if there is a chance of a major advance. "This is an invitation not to do the standard thing," he says.

Halliday, a particle physicist and former chairman of an *ad hoc* PPARC committee that looked into privatizing the management of Britain's royal observatories, says it is time that Britain won a Nobel prize for physics. (The last Briton to receive the physics prize was Sir Nevill Mott in 1977.)

He singles out two areas in which he feels that important breakthroughs, perhaps even a Nobel prize, are possible: gravitational waves — the bursts of waves in space produced by violent astrophysical events such as collapsing stars, predicted by Einstein's theory of general relativity — and the search for the missing 'dark matter'.

Members of research groups working on these two topics appear excited by Halliday's interest, although they are a little wary as they have not been PPARC's priorities in the past. Neil Spooner, a PPARC fellow at the University of Sheffield, which belongs to the UK Dark Matter Collaboration Project,



Halliday: keen to target resources on 'risky' projects where a Nobel prize might be won.

points out that PPARC's failure to maintain funding in previous years has set the group back, while groups overseas are investing much more heavily in dark-matter detectors.

The British dark-matter group is trying to establish whether weakly interacting massive particles (WIMPs) are a source of the missing mass. Effective detectors need to be massive and situated underground to exclude natural background radiation.

Spooner says PPARC's lack of interest has meant that the British group can afford only a 5-kg detector housed in a Yorkshire mine provided free of charge by the mine's owners. By contrast, a group at the University of California at Berkeley and Stanford University has access to a special cryogenic detector, and a group at the University of Rome is building a 250-kg detector.

PPARC's funding of British research into gravitational waves tells a similar story. The

University of Glasgow is a "world-leading" group, says Norna Robertson, reader in physics and astronomy. She says gravitational waves need long and highly sensitive detectors, ideally several kilometres in length, but Glasgow's detector is just 10 metres long.

The university is now building a part-PPARC-funded, 600-metre detector in Germany with the University of Hanover and the Max Planck Institute in Garching. But other groups have better equipment. A separate collaboration between Caltech and the Massachusetts Institute of Technology will result in two separate 4-km detectors at Hanford, in Washington state, and in Louisiana. And a French/Italian collaboration is building a 3-km detector near Pisa in Italy.

Halliday recognizes that both fields are highly competitive. But he believes it is still not too late for Britain's physicists to make significant progress, although creative thinking and a willingness to take intellectual risks are required. He says that a similar approach is needed for Britain's contribution at the European Laboratory for Particle Physics (CERN) in Geneva.

Until 2005, most of Britain's particle physics budget is committed to work at CERN. Indeed, two weeks ago, PPARC's governing council set aside £104 million (US\$173 million) for 16 British research groups to help build two detectors for the Large Hadron Collider (LHC), now under construction.

Halliday says he is committed to the LHC, and does not want to convey the impression of being an adversary. But he does not want British physicists to be "so busy building equipment" for the collider that they are unprepared for analysing "the new and exciting physics" expected to emerge once it begins to generate results.

Some physicists endorse his approach. Roger Cashmore, chairman of the University of Oxford's physics department, agrees that particle physicists in Britain will at some stage "need to move into analysis mode", but that this will be difficult while attention is focused on building the LHC detectors.

Others are more sceptical. John Ellis, a senior theoretical physicist at CERN, says it is hardly fair for Britain's physicists to "twiddle their thumbs" while the LHC detectors are being built, and then to "come in, do the analysis and walk off with the prize."

According to Halliday, "the [collider] is a major PPARC project. It will discover the Higgs particle, and the intellectual and experimental leaders will get Nobel prizes". He also accepts that most particle physics will be done at CERN for the next 15–20 years. "It is important for the future of particle physics that PPARC assists UK physicists to attain those roles."

Ehsan Masood

Britain opens up science advisory panel

[LONDON] A move by the British government to revive its senior advisory panel on research policy has been welcomed by science policy experts, who hope the revamped panel will improve on its previously moribund performance.

Margaret Beckett, president of the Board of Trade, announced last week that the Council for Science and Technology will be reconstituted with a large independent membership of 14 and a broad but clearly defined mandate. It will be chaired by Beckett, with Sir Robert May, the government's chief scientific adviser, serving as deputy chair.

The new mandate requires the council to advise the prime minister on strategic matters, such as the international benchmarking of UK research and the implications of public spending plans. The council will publish an annual report, and will "normally" publish its advice to the

government; the old council did not publish its work, and its members were not even publicly identified.

John Mulvey, director of the pressure group Save British Science, says his group welcomes the changes: "It looks likely to be a much more effective body than the previous council."

But others point out that the proposed reform falls short of a proposal made last year by the Dearing Commission on Higher Education for a new body, with an independent chair, to advise on research policy. They also suggest that a panel headed by Beckett may carry little weight outside her department.

In a separate move, the House of Commons Select Committee on Science and Technology, chaired by Michael Clark, is gathering evidence for an inquiry into the UK science advisory system which it plans to complete by the autumn.

Colin Macilwain

Dating in doubt as researcher is probed

[SAN DIEGO] A prominent geographer at Arizona State University (ASU), whose rock-surface dating method has been used internationally to age petroglyphs, as well as landforms at a nuclear waste dump site, is under scrutiny for possible scientific misconduct, according to officials.

Authorities are trying to determine whether Ronald I. Dorn added mixtures of charcoal and bituminous coal to samples to secure specific radiocarbon-dated ages of rock surfaces, or whether the two substances occur naturally on the surface material, called rock varnish.

Dorn, a tenured geography professor at ASU in Phoenix who is funded by the National Science Foundation (NSF), declines to discuss the inquiry. In an interview, he maintained that he and other researchers have shown that charcoal and bituminous coal are naturally present in rock varnish.

ASU and NSF officials also decline to comment on the inquiry, as do Dorn's attorneys in Washington DC, who have undertaken an aggressive legal defence, including threatening at least one journal in a bid to block publication of a potentially critical article.

But Nicholas A. Goodman, an attorney at the University of Arizona (UA) in Tucson, confirms that inquiries are under way into discoveries by scientists at the university that may raise questions about the findings of a number of Dorn's published articles. Goodman also says that the NSF's Office of Inspector General subpoenaed and received some sample residues of Dorn's that had been stored at a UA laboratory.

The inquiries are examining "the veracity of the findings of our researchers as compared to the findings of Dr Dorn," says Goodman. Because both ASU and UA are governed by the Arizona Board of Regents, the dispute pits sister academic institutions against each other. A scientist from Northern Arizona University in Flagstaff, governed by the same regent board, is also involved.

The NSF received the first allegation about Dorn's work in September 1996. But sources say that ASU's probe did not begin until early last year. ASU's scrutiny is now at the initial inquiry level, having not advanced to the point of a faculty committee of investigation, which under university guidelines could determine scientific misconduct, a source said.

In addition to the usual sensitivities of misconduct probes, this inquiry has an added twist: Dorn's father-in-law, Paul Sypherd, is the provost at the University of Arizona. Goodman says that the provost, second in command at UA, has nothing to do with any aspect of the scientific inquiry.



Rock of ages: the results of Dorn's dating techniques were used during a public debate in Portugal over plans to flood the Côa river valley (top), a move that threatened Stone Age carvings (below).

For several years, Dorn's methods have been of great interest — and some controversy — to a small group of scientists who try to determine the age of hard-to-date landforms. Because of this, the results of the probe are eagerly awaited in the discipline. If the inquiry were to find that the discrepancies were due to misconduct rather than error, this would also have wider implications because of the controversial nature of some of the sites Dorn has studied.

For instance, Dorn was involved in dating landforms at Nevada's Yucca Mountain, where US officials plan to bury highly radioactive nuclear waste. He also dated petroglyphs from Australia and from Portugal, where he analysed the age of rock art found in the Côa river valley. Originally intended to be filled as part of a reservoir, the valley was eventually deemed a park.

The UA team, along with scientists from three other institutions — the Lamont-

Doherty Earth Observatory at Columbia University in New York, Eidgenössische Technische Hochschule in Switzerland, and Northern Arizona University — are preparing to publish a critical paper in *Science* on some of Dorn's work. The scientists, some of whom have been co-authors with Dorn on previous papers, decline to discuss the substance of their article until it is published.

But the UA scientists — Warren Beck, Douglas Donahue, G. Burr and A. J. T. Jull — have twice previously presented abstracts of some aspects of their research that raise questions about Dorn's dating method. The abstracts, which are virtually identical, were presented last year at the Australasian Archaeometry Conference in Sydney and the International Radiocarbon Conference at Groningen in the Netherlands.

The abstracts say that "unusual and potentially very important observations" were found when examining samples of material Dorn submitted to date petroglyphs in the petrified forest region of north-eastern Arizona. Those petroglyphs, the work of early Anasazi, an ancient American Indian people, were discovered by Ekkehart Malotki of Northern Arizona University, who had sought Dorn's assistance in dating the drawings. The NSF-funded accelerator mass spectrometry laboratory at UA, where the authors of the abstracts work, carried out the radiocarbon analyses on behalf of Malotki and Dorn.

"Microscopic examination of samples from these petroglyphs, collected by and submitted to our laboratory by Dr Dorn, showed that the samples contained two types of black, carbon-rich materials with distinctly different visual properties," the abstract says. "Detailed examination of these particles revealed that one type strongly resembles finely ground bituminous coal, whereas

the other strongly resembles ground pyrolyzed wood."

Because these two substances have "widely differing radiocarbon ages", the abstract says, "the radiocarbon age of the entire sample would yield results which are, at best, ambiguous." The UA scientists added: "We have been unable to find either of these two types of carbonaceous material in equivalent samples of these same petroglyphs when subsequently resampled independently."

Asked about these results, Dorn says he was not at either conference, requested copies of the papers and noted that other scientists also have found such substances when studying rock varnish. As examples, he cites an article published in 1961 by a Russian scientist, N. N. Karlov, who is said to have observed "a coal-black mineral substance" trapped under "desert lacquer" in the Black Sea region. He also cites a 1986 doctoral thesis by a researcher from Texas Tech University in Lubbock, who studied petroglyph surfaces in New Mexico and West Texas.

"I don't understand what all the fuss is about," says Dorn, who adds that he himself began publishing papers in 1996 questioning the interpretation of his own dating in previous articles. "The existence of the different types of material with different ages means



Yucca Mountain: data on landforms were used in the debate on a proposed nuclear waste site.

that prior results are very problematic."

But while Dorn says the content of rock varnish has been the subject of collegial scientific discussions for some time, his attorneys were involved last year in trying to prevent the UA scientists from publishing some of their findings, which were made about the same time that Dorn began seeking publication of his self-critical articles.

Stephen C. Porter, a University of Washington geology professor who is editor of the journal *Quaternary Research*, says he received a strongly worded letter last November from Dorn's attorney, who was seeking to block

publication of an article after the UA scientists had offered Dorn co-authorship of it.

"The letter was rather intimidating," said Porter. "In 25 years as an editor, I've never experienced someone trying to prohibit publication prior to the review process." Dorn has since asked whether a full manuscript was submitted, says Porter, who adds that he never received the UA manuscript.

Porter identifies Dorn's attorney as Joseph N. Onek, who was Robert Gallo's attorney during his fight with the National Institutes of Health over a major HIV controversy. Onek has since moved to the US Department of Justice, according to a recent telephone message at his former law office. His science cases have been referred to another attorney, Robert P. Charrow. In a brief interview, Charrow made no comment, declining to acknowledge that he represented Dorn.

The forthcoming *Science* paper is expected to provide a new chapter in the debate, possibly determining the next step in Dorn's career, which has been studded with important awards, including the Geological Society of America's Kirk Bryan Award, of which he was a joint winner in 1986, and later an NSF Presidential Young Investigator Award.

Rex Dalton

Japan's genome programme goes ahead, with protein analysis

[TOKYO] Japan's largest and most advanced genome research programme will begin at the end of this year, despite an earlier dispute between its scientific advisers about whether it should become deeply involved in protein analysis.

The Genome Frontier Programme, now renamed the Biomolecular Research Programme, is being backed by the Institute of Physical and Chemical Research (RIKEN) and the Science and Technology Agency. It represents one of the latest efforts by the Japanese government to promote human genome research.

Support for genome research has grown dramatically in Japan over the past few years. The overall budget for the 1998 fiscal year, beginning 1 April, is one of the tightest for many years. But genome-related research projects across all science-related ministries and agencies saw a generous increase of more than double last year's figures (see *Nature* 391, 111; 1998).

But when details of the project — designed to incorporate both DNA sequencing and protein analysis — were developed last year, some leading genome researchers acting as advisers were said to be opposed to proposals for the protein-related research project (see *Nature* 389, 772; 1997). They argued that only DNA base sequencing should be carried out, since genome research in Japan was insufficiently advanced, and the

protein work would be premature.

But the government has now decided that the RIKEN programme will be carried out as originally planned. It will have three parts: DNA base sequencing using full-length mouse complementary DNAs, the analysis of the human genome using chromosome 21, and the analysis of protein function and structure.

Construction of the Biomolecular Research Centre, planned to be built in Yokohama, near Tokyo, is due to be completed in 2000, and there are plans to recruit at least 200 researchers. RIKEN also hopes to recruit 20 to 40 external researchers to participate in the project through a collaborative research system, and intends to accept proposals from overseas researchers in related fields.

The first five years of the 15-year project have recently been approved by the government. A sum of ¥4 billion (US\$31 million) will be spent in the programme's first year, and a total of ¥50 billion is planned to be invested over the five-year period.

Sequencing of full-length cDNA will be carried out using the latest automated high-speed sequencer developed by RIKEN, called RISA, for producing a genetic 'encyclopaedia' of gene and protein functions. The structure and function of proteins will be analysed using nuclear magnetic resonance (NMR) at a proposed

'NMR Park' (see *Nature* 381, 105; 1996) that has been incorporated into the new centre.

Part of the protein-related research will also be carried out at SPring-8, the world's largest third-generation synchrotron radiation facility, where RIKEN has two beamlines. The advanced features of the 8-GeV synchrotron, which began operation last October, are said to make it particularly useful for analysing protein structures.

Some leading genomic researchers claim the project is too ambitious because Japan is "several years behind the West" in genome research. Many also doubt the capabilities of RIKEN's new high-speed DNA sequencer, which is still under development.

But Akira Kira, the executive director of RIKEN, says that, although the technical aspects of the new sequencer have not yet been established, "we are optimistic about its potential". According to one project leader from RIKEN, RISA will help to minimize costs and could potentially carry out DNA sequencing at a much higher rate than commercially available models.

Akiyoshi Wada, former dean of science at Tokyo University and the director of Sagami Chemical Centre, says the project could potentially pave the way for Japan to join the front rank of international genomics research. But he warns that it will require careful management to fulfil its aims.

Asako Saegusa

China's science reforms may be less than revolutionary

[TOKYO] Despite drastic plans to streamline the Chinese government, announced at the People's Congress this month, and clear signals that technological innovation has become an important political priority, the administration of basic science in China is unlikely to change significantly, according to some observers.

Plans to reduce the number of state ministries and commissions from 40 to 29 were approved by the congress on 10 March. The State Science and Technology Commission, which is responsible for science and technology policy, will be renamed the Ministry of Science and Technology. The State Education Commission, which oversees university research, will also become a ministry.

But in practice the running of these organizations and their responsibilities are unlikely to alter, says a member of the Chinese Academy of Sciences, which is optimistic about the general possibilities that the reforms are opening up.

The most dramatic changes will take place in ministries related to industry. The Ministry of Chemical Industry will be downgraded to the National Bureau of Petroleum and Chemical Industry by merging with two government-run petroleum companies.

Early warning system for rare diseases planned

[STRASBOURG] The European Parliament last week approved the European Commission's proposal for a five-year research programme on 'rare diseases', a term that covers more than 5,000 conditions, including Creutzfeldt-Jakob disease.

But funding details are not yet clear. The commission has specified a budget of ECU1.2 million (US\$1.3 million) for the first year only, whereas the parliament has called for a total of ECU14 million over five years.

European Union research activities will start in 1999. They will include a free database, an early warning system to detect and assess rare diseases and the creation of rapid response teams.

Ministry kept BSE data under wraps for years

[LONDON] The UK Ministry of Agriculture consistently refused access to data on the epidemiology of bovine spongiform encephalopathy (BSE), according to a leading British epidemiologist. Roy Anderson, professor of zoology at the University of Oxford, told the BSE inquiry

that he was denied access to data despite making several formal approaches between 1989 and 1991.

Anderson said he was eventually given access in 1996 after seeking the help of a former colleague, the chief scientist, Sir Robert May. But he told the inquiry that getting access to the BSE database "continued to be a struggle" in 1997.

Anderson said that, if the ministry had made the data available in 1989, researchers would have discovered much earlier that infections were continuing, and the size of the epidemic could have been significantly smaller.

EU considers car makers' offer to cut emissions

[LONDON] Ritt Bjerregaard, the European Union's environment commissioner, has responded cautiously to an offer of voluntary cuts to carbon dioxide emissions from Europe's car makers. The manufacturers have offered to reduce emissions by 25 per cent from 1995 levels. Bjerregaard said she would "study the offer with an open mind".

Europe's environment ministers will meet on 23 March to discuss the offer. The proposed cuts exceed those offered by the industry last year. But in return the industry wants assurances that taxes on diesel cars will not be introduced.

India and Sweden line up joint research projects

[NEW DELHI] Sweden and India have decided to boost their 19-year science cooperation by launching a programme of joint research and development in food processing, environment and energy technology. At a two-day meeting in New Delhi last week, Indian scientists and a 20-member Swedish delegation identified 30 projects for funding.

Sweden proposes to contribute US\$5 million over three years. Sweden already spends nearly \$860,000 a year in India on joint projects in biotechnology and renewable energy.

Tobacco company found cancer link in 1970

[LONDON] A cigarette company memorandum from 1970 acknowledging "beyond all reasonable doubt" a link between smoking and lung cancer has been uncovered in the United States. The four-page report was sent to the managing director of Gallaher Ltd, a tobacco company being sued by 50 smokers, by the general manager of the company's research department. It is a review of the 'Auerbach experiment' on the effects of cigarette smoke on beagles.

The report said: "One of the striking features of the Auerbach experiment was that practically every dog which smoked suffered

significantly from the effects of the smoke either in terms of irritation and bronchitis, pre-cancerous changes, or cancer. The results of the research appear to us to remove the controversy regarding the causation of the majority of human lung cancer."

Gore launches idea for live pictures from space

[WASHINGTON] US Vice-President Al Gore proposed last week that the NASA space agency should launch a satellite by 2000 that would return live images of the Earth from space, 24 hours a day. Pictures from the satellite, which Gore dubbed Triana, would be downloaded from the Internet.

NASA plans to ask the educational, scientific and commerce communities for ideas for a \$50 million satellite equipped with an eight-inch telescope and high-definition TV camera. But the project has been derided as a "visual circus" by Dana Rohrabacher (Republican, California), who chairs the NASA oversight subcommittee in the House of Representatives.

Controversial cancer treatment set for trials

[LONDON] The Italian government has set up a commission to oversee clinical trials of a controversial, non-toxic cancer treatment following intense public demand for the drug (see *Nature* 391, 217; 1998). Last week, 20,000 people took to the streets of Rome demanding access to the treatment, devised by Luigi di Bella, a doctor from Modena, northern Italy.

Three experts, chaired by Gordon McVie, director-general of the UK Cancer Research Campaign, will evaluate trials of the therapy, known as MDB (Multiterapia di Bella). Di Bella says treatments should concentrate on encouraging healthy cells to multiply rather than on destroying infected cells.

Astronomy will make Greenwich comeback

[LONDON] The London borough of Greenwich will once again become a focal point for the public appreciation of astronomy, following the decision by the Particle Physics and Astronomy Research Council (PPARC) to transfer some of the activities of the Royal Greenwich Observatory back to its original site. The move follows the decision to transfer much of the technical work of the observatory from Cambridge to a new UK Astronomy Technology Centre in Edinburgh. Other activities will be relocated to the National Maritime Museum in Greenwich. According to PPARC, the partnership with the museum "offers the opportunity of creating a flagship for UK astronomy".

Some issues for the biosafety protocol

Sir — The letter from Crompton *et al.*, describing themselves as “individuals involved in either the development or implementation of risk assessment frameworks for biotechnology” — regulations mandated by the Convention on Biological Diversity (CBD) — illustrates perfectly the emerging problems of those efforts (*Nature* **391**, 528; 1998).

The authors assert that biosafety regulations should involve “the overt consideration of socioeconomic issues before the environmental release or transboundary movement of genetically modified organisms” (defined narrowly as those crafted with molecular techniques of genetic manipulation). In other words, any field trial, anywhere, of a new, recombinant strain of *Rhizobium* or a new variety of transgenic maize would be subject to an evaluation of possible social and economic impact, anywhere else in the world.

The basic premise of the CBD-mandated biosafety protocol — that the most precise, molecular techniques of genetic manipulation deserve extra scrutiny because they confer on products incremental risk — is contrary to scientific consensus. As a leading article in *Nature* has said (**356**, 1–2; 1992), a broad scientific consensus holds that “the same physical and biological laws govern the response of organisms modified by modern molecular and cellular methods and those produced by classical methods [Therefore] no conceptual distinction exists between genetic modification of plants and

microorganisms by classical methods or by molecular techniques that modify DNA and transfer genes.” Triggers to government supervision should therefore focus on product traits that may be related to risk rather than on whether one or another technique of genetic manipulation was used.

Furthermore, a requirement to consider socioeconomic factors before the field testing of a recombinant organism would imply the following kinds of scenarios and mandatory assessments.

- *Any pest- or disease-resistant plant:* What would be the possible impact on workers, companies and countries manufacturing chemical pesticides?

- *New plant varieties engineered to produce plastics or other substances now derived from petroleum:* What would be the effects on the economies of the Persian Gulf states and other OPEC countries?

- *A new malaria or schistosomiasis vaccine delivered in an edible fruit:* How will national infrastructures absorb increases in population if the death rate from those highly prevalent diseases falls precipitously?

If such questions can ever be answered accurately, there is certainly little likelihood of doing so while products are at an early stage of testing. At any stage, such efforts would constitute major interdisciplinary research projects and would be highly vulnerable to the vagaries of value judgements. Moreover, such analyses have not been required for other technological innovations, from the internal combustion

engine and electric automobiles to semi-dwarf wheat varieties and seedless grapes.

Finally, there is the more basic issue of whether a commitment to free trade and markets makes such requirements appropriate or useful at any time.

Henry I. Miller

Hoover Institution,
Stanford, California, USA
e-mail: miller@hoover.stanford.edu

Charles J. Arntzen

Boyce Thompson Institute,
Ithaca, New York, USA

Roger N. Beachy

Division of Plant Biology,
Scripps Research Institute, USA

R. James Cook

Department of Plant Pathology,
Washington State University, USA

Susanne L. Huttner

University of California,
Systemwide Biotechnology Program,
Berkeley, California, USA

Donald Kennedy

Institute for International Studies,
Stanford University, USA

Calvin O. Qualset

Genetic Resources Conservation Program,
Division of Agriculture and Natural Resources,
University of California, USA

Peter H. Raven

Missouri Botanical Garden,
and Washington University, USA

Anne K. Vidaver

Department of Plant Pathology,
University of Nebraska, USA

A pawn in a conspiracy?

Sir — Although I appreciate the legitimate bioethical concerns of emerging neurogenetic and neuroimaging technology, I do not believe the slope is quite as slippery as Jean-Pierre Changeux and Denis Le Bihan suggest (see *Nature* **391**, 316; 1998). Neuroimaging hardly seems capable of “invasion of personal liberty, control of behaviour and brainwashing”, and the assertion that neuroimaging can “almost read people’s thoughts” is scientifically ridiculous and philosophically naive.

They paint a frightening picture of a post-apocalyptic Big Brother neuroscientist armed with pocket magnetic resonance imaging that works at a distance in real time on moving subjects engaged in God-knows-what cognitive and metabolic activity and meaningfully translates these signals into a universal language revealing the subjective thoughts of the individuals scanned, only then to control their minds

and behaviour. Such broadcasting of thoughts and mind control sound like the results of schizophrenic processes rather than bioethical enquiries. Of course, I’m probably just a pawn in the conspiracy.

George A. Mashour

Georgetown University School of Medicine,
3970 Reservoir Road, RB-W225,
Washington, DC 20007, USA

Early Alpine industry

Sir — One of the ‘great mysteries’ of Ötzi, the Stone Age man whose body was found in an Alpine glacier (*Nature* **391**, 318; 1997), was recently solved. A copper axe had been found close to the body, which was (as you indicated) dated as more than 5,000 years old. Copper tools were not yet known from such an early time in prehistory, so that some scientists believed that the body and the axe were found together only incidentally.

It was proved recently, however, that Ötzi lived not only contemporaneously with the axe, but also that he was probably involved in its manufacture. This was indicated by analysis of hair samples, which showed increased copper and arsenic content, comparable to concentrations found nowadays in the hair of people who work — under relatively primitive conditions — in the copper industry.

The Alpine region where Ötzi was found contains, indeed, some morphological remnants of pits where, apparently, copper ores were mined in prehistoric times. This makes the likelihood of a copper industry during Ötzi’s lifetime almost certain. The finding of the ‘Ice Man’ thus sheds new light on the development of a metal industry in prehistoric times.

A. J. (Tom) van Loon

Geocom, PO Box 336,
NL-6860 AH Oosterbeek,
The Netherlands
e-mail: tom.van.loon@wxs.nl

A checkpoint on the road to cancer

Terry L. Orr-Weaver and Robert A. Weinberg

Mutations that disrupt a cell-division checkpoint, thereby causing alterations in chromosome number, have been identified in cancer cells. The accompanying increase in mutability helps to explain how tumours acquire large numbers of mutant genes during their development.

For the past two decades, cancer researchers have grappled with a paradox. The appearance of a human tumour is the culmination of a complex, multi-step process. The rate-limiting steps seem to be mutations in a half-dozen or more cellular genes that, directly or indirectly, affect tumour cell proliferation. However, from calculations using the known mutation rate in non-germline cells ($\sim 10^{-7}$ per gene per cell generation), one can predict that so many mutant genes would never accumulate

in the genome of cell lineage during a human lifetime^{1,2}. Hence, tumour formation is mathematically impossible. In the immortal words of the Maine farmer, answering a visitor's query for directions, "Ye cahn't get theah from heah". But experiments described by Cahill *et al.*³ on page 300 of this issue provide a road map to solving this paradox. They show that control mechanisms that ensure the proper separation of chromosomes during cell division can be defective in colorectal cancer cells.

Lawrence Loeb, whose calculations indicated that the completion of human tumour progression is highly improbable, revisited his numbers and concluded that one critical parameter assumed in his initial calculations — the mutation rate — had been underestimated^{1,2}. Tumorigenesis can occur, he concluded, only if the genomes of pre-malignant cells are far more mutable than those of their normal counterparts. Such mutability would hasten each rate-limiting mutational step in tumour progression, thereby accelerating completion of the process as whole.

Still, calculations like these are at best well-informed surmises, because too many critical parameters cannot yet be measured and must, therefore, be assumed.

More compelling have been direct demonstrations of increased tumour cell mutability. Many colon cancer cells, for example, cannot effectively repair mismatched DNA bases. As a result, they sustain various types of mutation at a high rate, the most apparent of these being changes in their microsatellite DNA blocs — short stretches of DNA of very simple, repeating base sequence⁴. But these tumours account for only a portion of those surveyed.

Is acquired mutability typical of most tumours as they progress? Studies of hereditary non-polyposis colon cancer have provided a clue. Tumours of this type tend to have normal or quasi-normal complements of chromosomes (euploidy)^{5,6}, but show microsatellite instability⁷⁻⁹. Most colorectal tumours, however, show no microsatellite instability, but have abnormal chromosome number (aneuploidy) and loss of heterozygosity at many genetic loci. This suggests that generation of aneuploidy is an alternative mutagenic mechanism to defective mismatch repair for driving tumour progression.

Aneuploidy can result from improper allocation of the two chromatids of a chromosome to the two daughter cells during cell division (mitosis). Normally, a 'checkpoint' control monitors the proper assembly of the mitotic spindle — the cellular apparatus that will pull the chromatids apart. If chromosomes are not attached stably to the microtubules that form the spindle, this

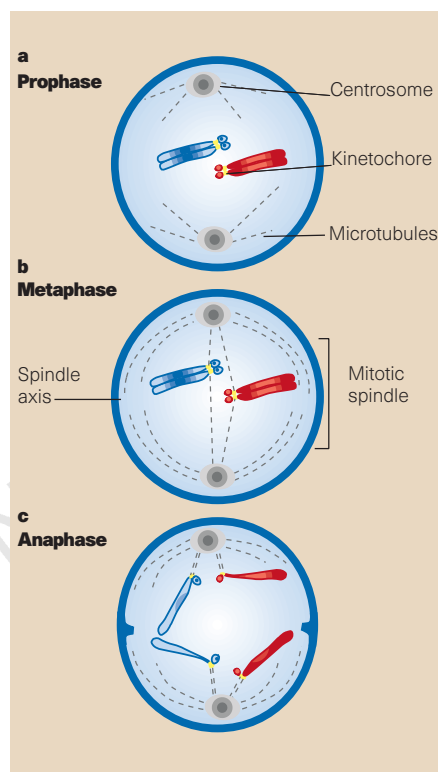


Figure 1 Normal separation of chromosomes during cell division (mitosis). a, Prophase. The spindle begins to form, and is visible as a series of microtubules (polymers of the protein tubulin), which are organized by the centrosomes found at either end of the cell. b, Metaphase. The spindle takes shape, and the chromosomes align themselves along the spindle axis, attached to the microtubules at a specialized region called the kinetochore. c, Anaphase. The chromosomes move along the shortening microtubules, kinetochores first, to opposite poles of the cell.

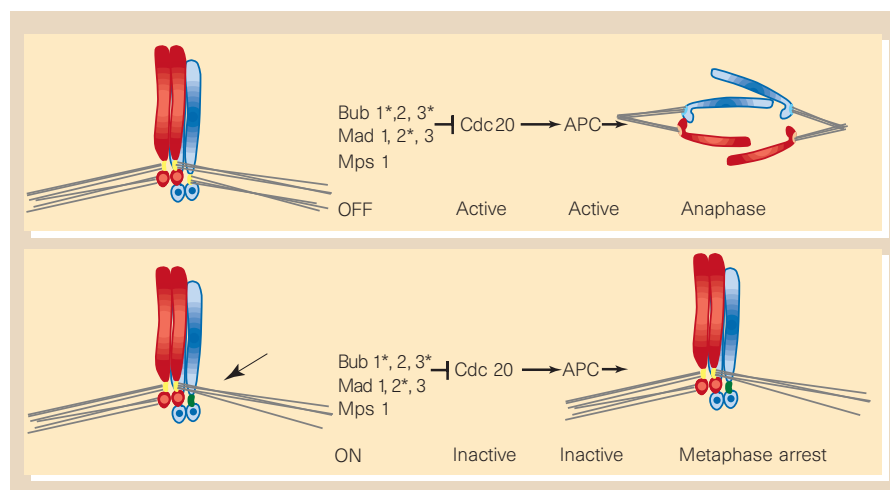


Figure 2 How assembly of the mitotic spindle is normally monitored to ensure chromosomal stability. In yeast, the Bub, Mad and Mps1 proteins respond to improper assembly of the spindle by blocking anaphase, apparently through inhibition of Cdc20, which would normally activate the cyclin destruction machinery (the anaphase-promoting complex, APC). The mammalian BUB1, BUB3 and MAD2 proteins bind to the kinetochore region, and it is possible that all of the proteins form a complex there. Normally, stable, bipolar attachment of the kinetochores leads to dissociation of BUB1 and MAD2 proteins from the kinetochore. But if the kinetochore is not attached to the microtubules, the BUB and MAD proteins remain attached, and arrest the cell in metaphase by delaying activation of the cyclin destruction machinery. Cahill *et al.*³ have found that this system is defective in tumour cells, leading to chromosomal instability (aneuploidy).

checkpoint control blocks the onset of anaphase (a stage of mitosis; Fig. 1). This block is lifted only when all of the chromosomes are stably attached (at a specialized chromosomal region known as the kinetochore) to the microtubules. Indeed, unattached kinetochores seem to emit a signal that prevents the onset of anaphase¹⁰. So failure of the spindle assembly checkpoint machinery leads rapidly to aneuploidy.

Some of the proteins that mediate this inhibitory signal in yeast cells have been identified. Mutations in their respective genes result in increased sensitivity to drugs that cause the microtubules to depolymerize and fall apart. This sensitivity arises because the drug-treated cells do not undergo mitotic arrest, indicating a non-functional checkpoint. Six genes have been recovered using this strategy — three *BUB* genes (budding uninhibited by benomyl) and three *MAD* genes (mitotic arrest-deficient)^{11,12}. A protein kinase (an enzyme that adds a phosphate group to its substrate), Mps1p, also seems to function in checkpoint signalling¹³.

The proteins that make up the spindle assembly checkpoint machinery seem to act directly at the kinetochore, to sense microtubule attachment and to send an arrest signal in its absence (Fig. 2). In mammalian cells, the *BUB1* and *MAD2* proteins are present at the centromere region in prophase of mitosis, but not after stable microtubule attachments are made at metaphase^{14,15}. *MAD2* has been found to bind CDC20; this protein activates the cyclin destruction machinery^{16,17}. This destruction machinery, also termed the anaphase-promoting complex (APC), degrades proteins that inhibit the separation of chromosomes, thereby permitting anaphase chromosome movement. So *MAD2* — or a complex of checkpoint proteins — inhibits the APC after it has sensed that the spindle attachments are defective.

Many aneuploid colorectal tumour cell lines show continuous chromosomal instability when grown in culture¹⁸. Cahill *et al.*³ now extend earlier experiments to reveal that this instability derives from defective control at the spindle assembly checkpoint. Moreover, they find mutant alleles of the human (*h*)*BUB1* gene in two of 19 colorectal tumour cell lines that show high rates of aneuploidy. By expressing the two mutant versions of the *hBUB1* gene in euploid cells, the authors can disrupt mitotic checkpoint control. Unexpectedly, however, only one of the two *hBUB1* gene copies present in the tumour-cell genomes described by Cahill *et al.* is mutant. Such heterozygosity might suggest that these mutant alleles act dominantly, as the authors propose, or it may indicate that a single, intact copy of the gene is inadequate for normal function.

Still, two swallows don't make a summer. The significance of these results will become

apparent only when more mutant *hBUB1* alleles are found, and when mutant alleles of other mitotic checkpoint genes are documented and functionally characterized.

Much of contemporary cancer research is motivated by the tenet that if mutant alleles of a gene are repeatedly detected in tumour cell genomes, then these alleles probably confer selective advantage on evolving tumour cell clones¹⁹. In the present instance, the mutations affect a gene that controls the stability of a cell's chromosome complement, leading, in turn, to aneuploidy. Tumour-cell aneuploidy has long been speculated to be causally involved in tumorigenesis, but its importance has not been demonstrable. The work of Cahill *et al.*³ represents the beginnings of a proof that acquired aneuploidy is a specific driving force in tumour progression, rather than a distracting epiphenomenon of this disease.

How does aneuploidy facilitate tumour progression? One attractive idea is that aneuploidy increases the rate at which tumour-suppressor genes are lost, through the loss of heterozygosity (which results in the conversion of genes having two dissimilar versions into identical versions). Elimination of these genes is known to cause deregulated cell proliferation. Perhaps most satisfying,

however, is that Loeb's prediction — that mutability is a general characteristic of tumour-cell genomes — has, after so long, taken on new life and credibility. □

Terry L. Orr-Weaver and Robert A. Weinberg are at the Whitehead Institute for Biomedical Research and the Department of Biology, MIT, Cambridge, Massachusetts 02142-1479, USA.

- Loeb, L. A. *Cancer Res.* **51**, 3075–3079 (1991).
- Loeb, L. A., Springgate, C. F. & Battula, N. *Cancer Res.* **34**, 2311–2321 (1974).
- Cahill, D. P. *et al.* *Nature* **392**, 300–303 (1998).
- Eshleman, J. R. & Markowitz, S. D. *Curr. Opin. Oncol.* **7**, 83–89 (1995).
- Kouri, M. *et al.* *Cancer* **65**, 1825–1829 (1990).
- Frei, J. V. *Cancer* **69**, 1108–1111 (1992).
- Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D. & Perucho, M. *Nature* **363**, 558–561 (1993).
- Thibodeau, S. N., Bren, G. & Schaid, D. *Science* **260**, 816–819 (1993).
- Aaltonen, L. A. *et al.* *Science* **260**, 812–816 (1993).
- Rieder, C. L., Cole, R. W., Khodjakov, A. & Sluder, G. *J. Cell Biol.* **130**, 941–948 (1995).
- Hoyt, M. A., Totis, L. & Roberts, B. T. *Cell* **66**, 507–517 (1991).
- Li, R. & Murray, A. *Cell* **66**, 519–531 (1991).
- Hardwick, K. G., Weiss, E., Luca, F. C., Winey, M. & Murray, A. W. *Science* **273**, 953–956 (1996).
- Taylor, S. S. & McKeon, F. *Cell* **89**, 727–735 (1997).
- Li, Y. & Benezra, R. *Science* **274**, 246–248 (1996).
- Hwang, L. H. *et al.* *Science* **279**, 1041–1044 (1998).
- Kim, S. H., Lin, D. P., Matsumoto, S., Kitazono, A. & Matsumoto, T. *Science* **279**, 1045–1047 (1998).
- Lengauer, C., Kinzler, K. W. & Vogelstein, B. *Nature* **386**, 623–626 (1997).
- Nowell, P. C. *Science* **194**, 23–28 (1976).

Geophysics

Finding chaos in abyssal hills

John A. Goff

Abyssal hills cover most of the Earth's sea floor. They are, in fact, the most common morphological feature on the Earth's surface, yet they are among the least understood. These elongate hills are created at mid-ocean-ridge spreading centres by faults which offset the volcanically created sea floor, and can be seen as both simple and complex (Fig. 1).

The simple view is often taken by physical modellers, who look upon oceanic rifting as a test case in which the dynamic processes of an active Earth are barely complicated by factors such as erosion or sedimentation. Modellers tend to speak a deterministic language, specifying their predictions of faults, for example, in terms of regularly spaced fault intervals and uniform fault offsets^{1,2}. Observationalists, by contrast, tend to point to the random, chaotic-looking behaviour of the abyssal-hill fabric, and naturally choose the mathematical language of statistics to quantify their observations^{3,4}.

So observations and predictions have not been directly comparable, and this is the problem addressed by Buck and Poliakov in their paper on page 272 of this issue⁵. They have developed a numerical physical model of mid-ocean-ridge faulting that produces a chaotic-looking morphology very like that

of abyssal hills observed on the sea floor.

Statistical descriptions of natural phenomena have been around for some time, but such applications reached truly phenomenal proportions following publication of Benoit Mandelbrot's *The Fractal Geometry of Nature*⁶ in 1983. Although the mathematics described by Mandelbrot were not new (Hausdorff had much earlier explored the concept of fractional dimensions⁷), his book coincided with the rapid emergence of computer graphics. The resulting images were extraordinary — truly realistic representations of landscapes, seascapes, trees and other natural phenomena, as well as unreal, psychedelic drawings created from simple mathematical abstractions. Such graphics had an immediate impact on popular culture (the producers of the film *Star Trek: The Search for Spock* went so far as to create an entire world with fractal imaging techniques), as well as on education, with students scrambling to learn more about the equations that governed such 'cool' pictures.

Scientists also jumped on the bandwagon, wanting to understand how this 'new', fractal way of thinking about nature might bear on their particular speciality. Such enquiries, however, eventually lost a bit of

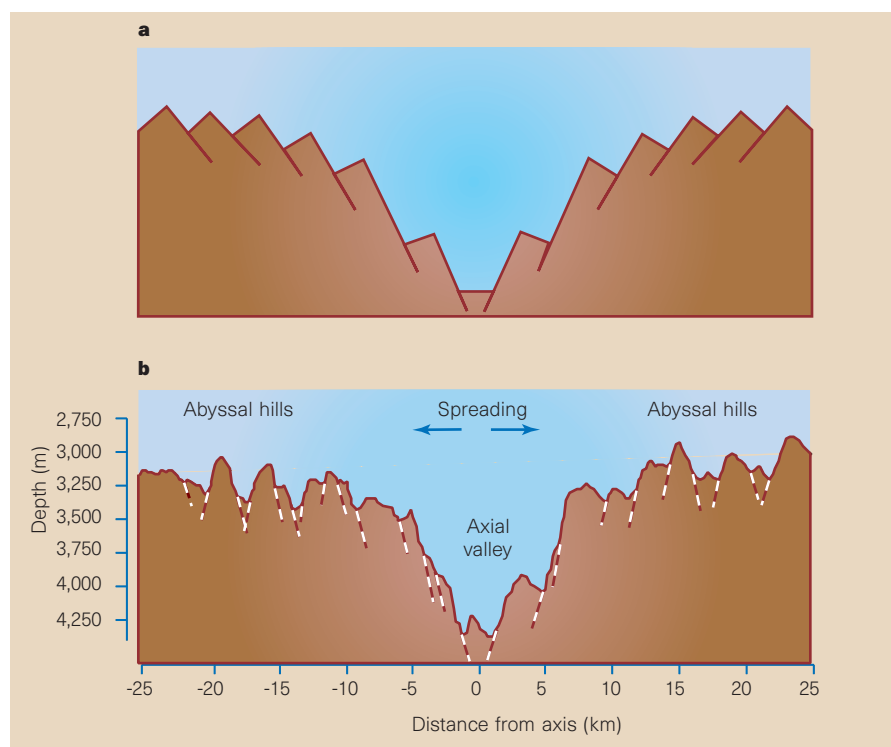


Figure 1 Two views of the formation of abyssal hills. **a**, An idealized representation of faulting at a mid-ocean-ridge rift axis with an axial valley. The idealized faulting does not reproduce the chaotic-looking nature of abyssal hills, as shown in **b**. **b**, Bathymetric profile across the southeast Indian Ridge near 115° E, typical of mid-ocean ridges with slow to medium spreading rates. Hypothetical fault locations are shown by dashed lines. Faults are driven by spreading localized at the mid-ocean-ridge axis, leading to the formation of an axial valley and, eventually, abyssal hills. Buck and Poliakov's numerical faulting model⁵, discussed here, reproduces the chaotic appearance of abyssal hills, thereby increasing our ability to predict the behaviour of such complex systems from physical principles.

steam. After too many measurements of the fractal dimension of this or the fractal dimension of that, a collective 'Now what?' seemed to rise from the scientific community. What does the fractal dimension, or any other statistical measurement for that matter, really mean? What do these descriptions tell us about the physics, geology or biology of the process being described? There is a gulf between merely describing the statistical behaviour of a natural phenomenon (which can also include properties such as mean, variance and correlation, as well as fractal dimension) and understanding the physical principles which lead to such behaviour, especially as one moves beyond the quantum world into the macroscopic.

That gulf has been bridged in select circumstances. For example, in 1948 von Kármán published an elegant treatise on the statistical theory of turbulence⁸, in which he used physical principles to infer how velocity cells cascade to ever smaller cells. From such arguments he derived the well-known 'two-thirds law' which characterizes the power law exponent of the spectral density function (a statistical description) for a turbulent velocity field. The power law exponent is, in essence, a proxy for the fractal dimension.

Von Kármán's work was one of the first

successful attempts to connect a physical description of a chaotic, macroscopic natural phenomenon (turbulence) with a quantitative statistical characterization of the same phenomenon. It also led to a generic descriptive statistical model that has been successfully used to quantitatively characterize a wide variety of completely unrelated geophysical phenomena, including abyssal hills².

Although it is perhaps the simplest example of faulting on the Earth, faulting at mid-ocean ridges is still too complex a system to derive statistical character from physical principles. Rather, these systems must be investigated numerically with models that incorporate as much of the relevant physics as possible. Buck and Poliakov's paper⁵ is the latest product of this line of research. Each new study of this kind has the advantage of increased computer power over previous ones, and a few years ago it would have been impractical to attempt to numerically model such a large and diverse faulting system as that which gives rise to abyssal-hill morphology. But Buck and Poliakov have also added a simple but critical innovation that allows for the spontaneous development of faults when a failure criterion is met; that is, when the stresses generated by plate



100 YEARS AGO

In this brochure, Mr Moxly expresses his dissatisfaction with the theory of the tides, as ordinarily accepted, and submits an alternative explanation. Although he opposes the views generally held, he does not exhibit that spirit of antagonism and adopt the language of abuse that too frequently disfigures the writings of those who dissent from authoritative teaching. ... The author considers that the tides are heaved up by the earth's gravity. The differential attraction of the sun and moon simply gives an opportunity for the earth's gravity to display itself in this manner. This action is illustrated by reference to a football. When the leathern covering is injured, or a seam gives way, the inner india-rubber case bulges out through the opening in the outer cover. "The pressure of the outer case had been removed from one region of the ball, and the pressure of the part which remained did the rest. This, I take it, is exactly how the pressure of the earth's gravity produces the tide." We are all prepared to admit with the author, that the tide-raising force is directly opposed to the action of the earth's gravity, though we might not adopt his phraseology. But another elementary proposition shows that the tide-raising force varies inversely as the cube of the distance from the disturbing body, and we fail to derive this from the football illustration ...

From *Nature* 17 March 1898.

50 YEARS AGO

For two hundred years, popular writers have attacked traditional religious beliefs in the name of science and pretty well undermined the faith of the plain man. Now, however, the sceptical attitude is turned against science itself, at a moment when, thanks to recent developments of physical theory, science appears vulnerable intellectually and, thanks to recent developments of military and political technique, still more vulnerable morally. Modern men, finding themselves with nothing to put their faith in, fill the spiritual vacuum with bogus religion based on bogus science, like Nazism and Communism. Even the more modest alternative, the assurance that all problems are technological and that the salvation of mankind depends on bigger and better gadgets, is not entirely satisfactory.

From *Nature* 20 March 1948.

spreading overcome the internal strength of oceanic crust. So rather than impose a regular fault spacing, this method allows the numerical model to find its own geometry, and the results look satisfyingly chaotic and realistic.

Buck and Poliakov have taken an initial step in comparing the statistical character of their model runs with the observed statistical character of abyssal hills. The results are promising, implying that they have accounted for the primary influences — such as crustal rheology, temperature, strain rate and cohesion strength — on the formation of abyssal hills, at least at ridges with slower spreading rates. But there is still much to be done. Abyssal hills vary widely over the Earth's ocean floors, and depend on the complex interplay of such factors as spreading

rate, ridge segmentation and axial morphology⁹. Detailed modelling along the lines developed by Buck and Poliakov will be required to fully understand these relationships. □

John A. Goff is at the University of Texas Institute for Geophysics, 4412 Spicewood Springs Road, Building 600, Austin, Texas 78759, USA.

e-mail: goff@utig.utexas.edu

1. Shaw, P. R. & Lin, J. J. *Geophys. Res.* **98**, 21839–21851 (1993).
2. Shaw, W. J. & Lin, J. J. *Geophys. Res.* **101**, 17977–17993 (1996).
3. Fox, C. G. & Hayes, D. E. *Rev. Geophys.* **23**, 1–48 (1985).
4. Goff, J. A. & Jordan, T. H. *J. Geophys. Res.* **93**, 13589–13608 (1988).
5. Buck, W. R. & Poliakov, A. N. B. *Nature* **392**, 272–275 (1998).
6. Mandelbrot, B. B. *The Fractal Geometry of Nature* (Freeman, New York, 1983).
7. Hausdorff, F. *Mathematische Annalen* **79**, 157–179 (1919).
8. von Kármán, T. *J. Mar. Res.* **7**, 252–264 (1948).
9. Goff, J. A., Ma, Y., Shah, A., Cochran, J. R. & Sempéré, J.-C. *J. Geophys. Res.* **102**, 15521–15534 (1997).

Malaria transmission

Has the ignition key been found?

Richard Carter and Lisa Ranford-Cartwright

Malaria parasites alternate between the tissues and blood circulation of a vertebrate host (for example, man) and those of an invertebrate host — mosquitoes of the genus *Anopheles* in the case of human malaria. The parasites are transmitted from humans to mosquitoes as male and female gametocytes, which are enclosed within the human red blood cell (Fig. 1). For successful transmission of parasites, the gametocytes must escape from their host red cells, transform into male or female gametes (gametogenesis) and fertilize, within 10 to 15 minutes of being taken up in a mosquito blood meal. This sequence of events must be reliably initiated in the mosquito blood, but totally suppressed within the human circulation. The key to how this is done can be found on page 289 of this issue¹, where Oliver Billker and colleagues describe the crucial factor that triggers gametogenesis of malaria parasites in mosquitoes.

The events of gametogenesis were classically termed 'exflagellation', from the formation of the vigorously motile male gametes out of the previously quiescent male gametocytes. This visually striking phenomenon has fascinated malariologists since the discovery of parasites themselves by Alphonse Laveran while, as it happens, observing the very process of exflagellation in malarious human blood². Laveran's discovery was possible because exflagellation occurs not only in mosquitoes, but also spontaneously on exposure of gametocyte-infected blood to air, and it is readily seen in a drop of gametocyte-containing blood viewed on a glass slide under a microscope.

Thus, it seemed self-evident that exposure to ambient atmospheric conditions — as when a few microlitres of blood are ingest-

ed by a mosquito — must involve all the signals necessary to initiate gametogenesis. Indeed, experiments done in the 1950s and 1970s seemed to show precisely this^{3–6}. The mechanism uncovered consisted of a bicarbonate-ion-dependent rise in pH, from a non-permissive pH of 7.4 in the host circulation, to a fully permissive pH between 8.0 and 8.3. This rise in pH is due to a fall in the concentration of carbonic acid, achieved by the loss of CO₂ from the blood on exposure to air. Temperature was also crucial. Whereas

at host body temperature exflagellation was completely suppressed, a fall by as little as 2–3 °C below body temperature was fully permissive⁷.

One fact, however, did not fit. The pH of mosquito blood meals rarely (if ever) rises as high as even pH 7.8 — the threshold for bicarbonate-ion-dependent induction of exflagellation^{4,6}. Then, in 1976, Mary Nijhout discovered another factor that acts independently of bicarbonate ions. By feeding gametocytes to mosquitoes in a simple glucose/saline buffered to pH 7.4, but without bicarbonate (a solution that cannot activate gametogenesis), she showed that exflagellation was rapidly initiated within the mosquito midgut. The blood meal contained a potent mosquito-derived factor that triggered gametogenesis, and seemed to function as an 'ignition key' for the mosquito phase of the malaria parasites' life cycle. She named⁸ this 'mosquito exflagellation factor', or MEF.

Subsequent (unpublished) characterization by Ray Beach and Nijhout revealed that MEF was organic in nature, heat stable, probably contained double-bonded ring structures, and had a relative molecular mass of less than 500 (ref. 7). There were no further reports until last year, when a heat-stable, hydrophilic molecule that had all the biological characteristics of MEF was purified chromatographically from the gut lumen and heads of unfed female *Anopheles*. The molecule was characterized as a small, negatively charged chromophore⁹. Now, in the series of investigations reported by Billker and colleagues¹, the chemical identity of

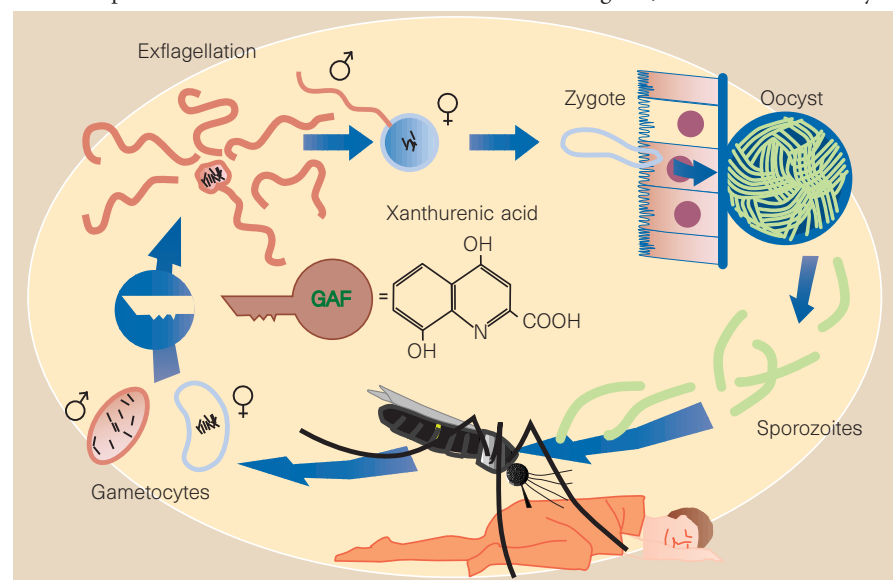


Figure 1 The malaria cycle. A female *Anopheles* mosquito takes a blood meal on a malaria-infected human. Male and female malaria gametocytes are immediately triggered to transform into gametes and to fertilize in the midgut of the mosquito. Billker *et al.*¹ have found that xanthurenic acid (a by-product in the synthesis of eye pigment in mosquitoes) is present in the mosquito blood meal. It seems to represent the crucial 'gametocyte activating factor' (GAF) that initiates gametogenesis and the subsequent development of the parasites in the mosquito vector. Eventually, sporozoites are produced — infective stages that sequester in the mosquito salivary glands and are injected into another human host to continue the cycle of malaria transmission.

MEF is convincingly demonstrated. Renaming it 'gametocyte activating factor' (GAF), the authors show it to be a single molecular entity — xanthurenic acid (Fig. 1).

The properties of xanthurenic acid, a simple two-ring structure derived from tryptophan, conform exactly to those described⁸ for MEF. Unexpectedly, in its capacity as MEF/GAF, xanthurenic acid is, after all, present in the blood of mammals¹⁰. But it is found only at concentrations that are below the threshold necessary to induce exflagellation. Importantly, as had been shown⁸ for MEF, Billker *et al.* found GAF activity not only in the midguts of mosquitoes, but also in head tissues¹. Xanthurenic acid is a by-product of the pathway that leads to the synthesis of ommochromes in the eye pigment of insects such as mosquitoes and the fruitfly *Drosophila*¹¹. The authors exploited this fact, as important proof that MEF, GAF and xanthurenic acid are one and the same, by showing that GAF activity is present in wild-type *Drosophila* but is absent from *Drosophila* eye-colour mutants defective in this metabolic pathway.

What remains to be investigated, and where might this discovery lead in relation to human disease? A crucial experiment will be to test the infectivity of human malaria parasites to eye-colour mutants of *Anopheles* mosquitoes that cannot produce xanthurenic acid¹². Should they be refractory to infection with human malaria parasites, the role of xanthurenic acid as the 'ignition key'

to transmission of malaria through mosquitoes will be effectively proved. In *Drosophila*, the genes for eye-pigment mutations are well characterized, and seem to confer a moderately small loss of fitness, at least in laboratory colonies. Their equivalent in *Anopheles* mosquitoes could represent the first identified genes for refractoriness to malaria in the vector¹³. Such a finding would suggest the intriguing idea of controlling the transmission of malaria by manipulations involving *Anopheles* eye-colour genes. □

Richard Carter and Lisa Ranford-Cartwright are at the Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK.

e-mail: r.carter@ed.ac.uk

1. Billker, O. *et al.* *Nature* **392**, 289–292 (1998).
2. Laveran, A. *Bull. Acad. Natl. Med. (Paris)* **9**, 1235 (1880).
3. Bishop, A. & McConnachie, E. W. *Parasitology* **46**, 192–215 (1956).
4. Bishop, A. & McConnachie, E. W. *Parasitology* **50**, 431–448 (1960).
5. Micks, D. W., de Caires, P. F. & Franco, L. B. *Am. J. Hyg.* **48**, 182–190 (1948).
6. Nijhout, M. & Carter, R. *Parasitology* **76**, 39–53 (1978).
7. Carter, R. & Graves, P. M. in *Malaria: Principles and Practice of Malariaology* (eds Wernsdorfer, W. H. & McGregor, I. A.) 253–306 (Churchill Livingstone, Edinburgh, 1988).
8. Nijhout, M. M. *Exp. Parasitol.* **48**, 75–80 (1979).
9. García, G. E., Wirtz, R. A. & Rosenberg, R. *Mol. Biochem. Parasitol.* **88**, 127–135 (1997).
10. Williams, S. A., Monti, J. A., Boots, L. R. & Cornwell, P. E. *Am. J. Clin. Nutr.* **40**, 159–167 (1984).
11. Kayser, H. in *Comprehensive Insect Physiology, Biochemistry and Pharmacology* Vol. 10 (eds Kerkut, G. A. & Gilbert, L. I.) 365–416 (Pergamon, Oxford, 1985).
12. Beard, C. B., Benedict, M. Q., Primus, J. P., Finnerty, V. & Collins, F. H. *J. Hered.* **86**, 375–380 (1995).
13. Besansky, N. J. *et al.* *Insect Mol. Biol.* **4**, 217–231 (1995).

Galaxy formation

Clumps and bumps on the road

Melinda L. Weil

Astronomers who probe the mysteries of galaxy formation traverse a bumpy road. Some galaxies are easy to distinguish by shape: there are spheroidal ellipticals, and spirals, which contain a bulge at the centre of a disk of stars in which spiral arms curl. But how do these different types form? Does one type evolve into the other? On page 253 of this issue¹, Masafumi Noguchi describes a simple simulation, in which a spiral galaxy forms from a single uniform sphere that first fragments into many large stellar clumps.

Sixty years ago, Hubble classified regular galaxies according to morphology. Figure 1 shows a progression from round to flattened ellipticals through the lens-shaped S0s to spirals, which are ordered according to the size of their central bulges and the tightness of their spiral arms. Hubble's diagram forks to distinguish between galaxies with and without bars. It was thought to be a map of evolutionary sequence, and so ellipticals were called 'early type' galaxies and spirals 'late type'.

When it became apparent that this simple scheme was not viable, the map's mirror

image was presented as an illustration of how two spiral galaxies might merge to form a spheroidal elliptical on the handle². In this scheme, spirals can be considered the 'early' galaxies that form the 'later' ellipticals. And, indeed, we see some galaxies that seem to be

creating ellipticals by merging³.

Numerical simulations of merging stellar disks can produce remnants that look like elliptical galaxies^{4,5}. Unfortunately, they have large cores of nearly constant density, unlike real ellipticals⁶. Merging spirals with large bulges alleviates most of the problem⁷ — but if the Universe refused to create elliptical galaxies without the assistance of major mergers, how did it create the spheroidal bulges?

The new simulation provides a possible answer. Noguchi treats a protogalaxy as a rapidly collapsing uniform sphere of dark matter and giant molecular gas clouds. Clouds are stuck together when they interpenetrate, regardless of their relative velocities, and entire clouds are turned into 'stellar' particles at a rate dependent on the gas density. Because each stellar particle really contains more than a million Suns, it is assumed to include massive stars that explode as supernovae and push neighbouring clouds about.

In the model, large clumps — of about a billion solar masses — are produced in the disk and orbit the centre of the galaxy. The clumps lose angular momentum by dynamical friction against the ambient medium, finally gathering in the centre to form a bulge. (Noguchi suggests that the bulge should contain stars with a range of ages and with correlations between their kinematics and abundances, as might the bulge of our own Galaxy.) Although this is promising, the model's robustness is not explored: clouds that touch don't necessarily stick together. Without the enforced stickiness, the clumps, which drive the evolution, would not form.

Is there observational evidence for Noguchi's clumps? Because the model collapses to a disk quickly, it best represents galaxies in the early Universe. Observations of such galaxies, with redshifts between four and two, uncover some that seem to be clumpy^{8,9}. But the Hubble Deep Field showed that, at later redshifts between about 1 and 0.3, symmetric spiral galaxies are rare whereas ellipticals are plentiful. (So we had better reverse Hubble's map again.) Whereas

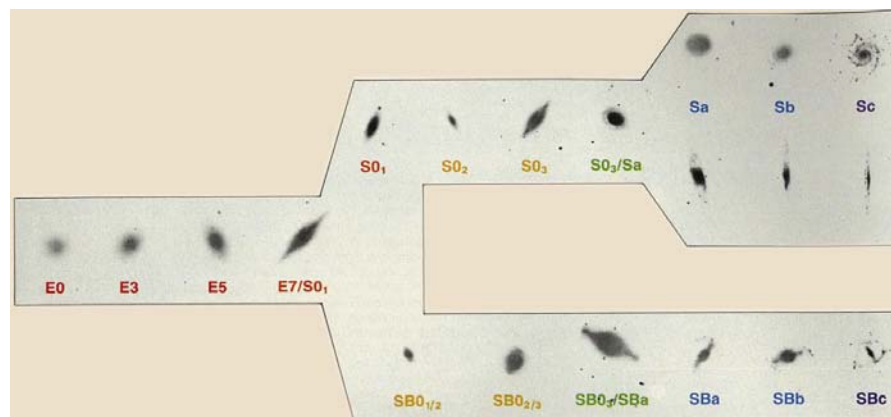


Figure 1 A Hubble 'tuning fork' diagram, classifying galaxy types. This version differs from Hubble's original in that lens-shaped galaxies (S0 and SB0) are located on the tines with the spirals (S and SB) rather than at the intersection of the tines and handle.

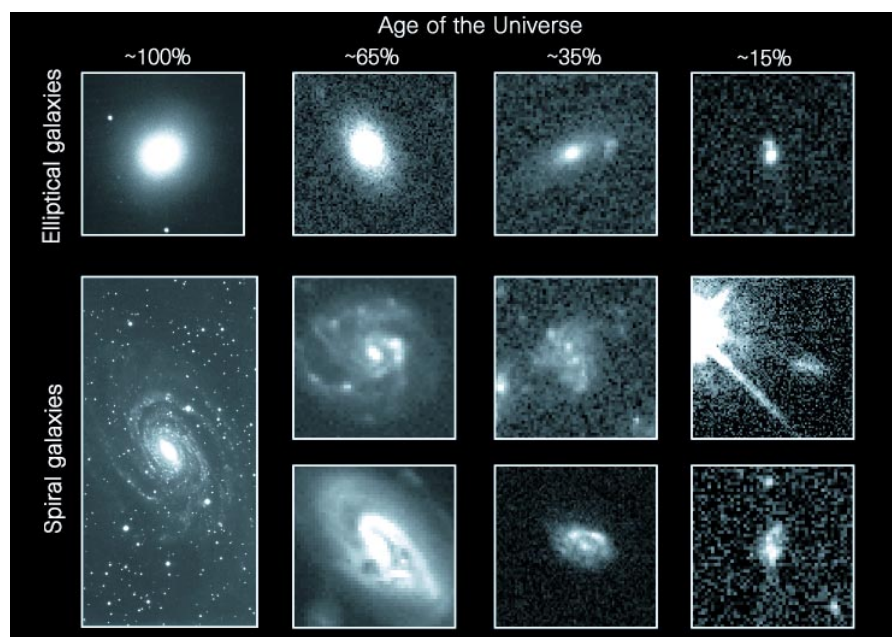


Figure 2 The evolution of elliptical and spiral galaxies. Whereas ellipticals are easily recognized even in the early Universe, the distinctive structure of today's spiral galaxies appears increasingly disturbed at earlier times.

probable spiral progenitors in the Hubble Deep Field are lumpy and twisted, at these redshifts Noguchi's disk is smooth.

However, in cosmological simulations that use our best guesses as to the state of the early Universe, galaxy formation proceeds hierarchically. Galaxies are built by mergers and fragile stellar disks are often destroyed when they combine to form hot, spheroidal objects at high redshifts¹⁰. Ummm, OK, let's invert our map again! Now we can account for ellipticals. But what road leads to the thin,

rotating spirals we see today?

Clumps to the rescue. In Noguchi's isolated galaxy, clumps help form disks with surface density profiles similar to those of real spirals. But it is difficult to determine whether the bulge-to-disk mass ratio in the model is representative of spirals, and whether a thin disk still exists at the final time.

How can we distinguish between Noguchi's model and a hierarchical picture, in which clumpiness could be produced by merging among small companion galaxies?

One difference is that, in Noguchi's model, the clumps should all rotate around a common centre, and observations may be able to detect such rotation.

What other phenomena might clumps cause? Noguchi suggests that they may be the seeds that produce active galactic nuclei, whose huge luminosities are attributed to gas being accreted onto a supermassive black hole. Mergers of orbiting clumps might feed active galactic nuclei, by inducing an inflow of gas. But neither in Noguchi's models nor in hierarchical models is the resolution high enough to explore this assertion in detail.

Noguchi's simulation is alluring in that it treats galaxy formation at a resolution not easily obtained by cosmological simulations. But such simple models both triumph and fail: because it does not rely on cosmology, the collapse need not be limited to a particular time in the Universe; but neither does it predict a cosmology that can be tested against the real Universe. Such models leave us with no map. □

Melinda L. Weil is in the Astronomy Department, Columbia University, 538 West 120th Street, New York, New York 10027, USA.

e-mail: mweil@astro.columbia.edu

1. Noguchi, M. *Nature* **392**, 253–256 (1998).
2. Schweizer, F. in *Dynamics and Interactions of Galaxies* (ed. Wielen, R.) 60–71 (Springer, Heidelberg, 1990).
3. Toomre, A. in *The Evolution of Galaxies and Stellar Populations* (eds Tinsley, B. M. & Larson, R. B.) 401–416 (Yale Univ. Press, New Haven, CT, 1977).
4. Barnes, J. E. *Astrophys. J.* **331**, 699–717 (1988).
5. Hernquist, L. *Astrophys. J.* **400**, 460–475 (1992).
6. Lauer, T. R. et al. *Astron. J.* **110**, 2622–2654 (1995).
7. Hernquist, L. *Astrophys. J.* **409**, 548–562 (1993).
8. Steidel, C. C., Giallisco, M., Dickinson, M. & Adelberger, K. L. *Astron. J.* **112**, 352–358 (1996).
9. Giallisco, M., Steidel, C. C. & Macchetto, F. D. *Astrophys. J.*

Ecology

Nipped in the bud

There are many examples of happy partnerships between plants and ants, in which, for instance, the plant gains protection from herbivores and in return the ants are housed. These symbiotic relationships can, however, be so one-sided that they verge on parasitism, and a striking example is described by Douglas Yu and Naomi Pierce in *Proceedings of the Royal Society of London B* (265, 375–382; 1998).

The plant involved is a member of the Boraginaceae family, *Cordia nodosa*, which grows to some 2 m in height and is a common constituent of lowland forest in southeast Amazonian Peru. The ant belongs to the genus *Allomerus*, and inhabits so-called domatia, specialized swellings in the stems of the host. Yu and Pierce found that almost 80% of the plants they sampled contained colonies of *Allomerus*.

The ant, however, manipulates the plant to its own ends by destroying the flower

buds and flowers while energetically protecting the leaves from attack by other insects. The ants don't eat the buds or flowers — they just wilfully chew off the reproductive structures in what Yu and Pierce refer to as acts of 'castration'. The result is ravaged buds, such as those shown here, and a dramatic reduction in fruit production. Conversely, senescence of the plant's branches is delayed, thereby providing more domatia for the ants. There appears to be no gain for the plant in this, just the pain of severely impaired reproductive capacity. Nor, apparently, does the plant retaliate in any way.

The question then is why, given that the host's competitiveness must be adversely affected, does this ant-plant relationship seem to persist and be widespread? Yu and Pierce think that one explanation may lie in the truly mutualistic relationship that *C. nodosa* has with three species of another genus of ant, *Azteca*. These species — the



good guys as far as the plant is concerned — inhabited only about one-tenth of the *C. nodosa* sampled. But these plants reproduce normally and may provide enough new recruits to keep the overall population stable.

Further twists lie in the observations that a few of the *C. nodosa* examined carried colonies of both genera of ant on separate trunks; the respective effects on plant reproduction were as expected from the main findings. More curiously, in fewer than 1% of cases the *Allomerus* ants left the buds alone, and the host fruited normally. That tantalizing finding defies explanation.

Tim Lincoln

Visual processing

How to know where to go

Roland Hengstenberg

When you move through a landscape containing objects at various distances, the images on your retinas move as you turn and change as you progress. Such image motions ('optic flow') are the inevitable consequence of locomotion, and they occur in any creature or robot with eyes. Wylie and colleagues have studied the processing of optic flow in the pigeon brain^{1,2} and, on page 278 of this issue¹, they describe in detail one component of that process.

On the one hand, movement of the retinal image interferes with vision — by producing motion blur, by displacing interesting objects from the area of best vision, or by seeming to tilt things that are upright. Most animals with good vision try to minimize these disturbances, at least temporarily, with gaze-stabilizing eye and head movements while they move around. On the other hand, however, optic flow can be a rich source of

information. Humans and animals exploit it to sense their own motion or to segment the visual scene into objects on the basis of motion parallax. We also use optic flow to estimate our distance from obstacles or targets, and to control other movements during locomotion.

Mathematically, any optic-flow field can be decomposed into a rotatory component (direction and rate of turning) and a translatory component (direction and speed of progression)³. Rotatory flow does not depend on the distance of the objects. In translatory flow, however, visual objects seem to move quickly when they are close by, gradually becoming slower the farther away they are. To make best use of the different types of information buried in the optic flow, it would be helpful to distinguish between the main components, and to work out their respective directions and rates separately. How is this done in the brain?

First of all, motion is not directly visible. At any given instant, the retinal image is represented by a two-dimensional excitation pattern over millions of photoreceptors. Motion must be computed locally in deeper layers of the visual nervous system. When a brightness change occurs, first in a small patch A of the visual field and a little later in an adjacent patch B, something has probably moved from A to B. The spatial arrangement of the two patches specifies the direction of motion. For every location in the visual field, motion is computed in different directions. This process goes on continuously, providing many simultaneous local motion signals all over and in all directions. Motion may be detected in retinal ganglion cells⁴, the primary visual cortex⁴, the accessory optic system⁴, or, for invertebrates, in distinct areas of the brain⁵. But the cellular mechanisms remain to be clarified⁵.

Second, the components of optic flow are not explicitly represented in the jumble of small field motion signals that are generated in the brain by local motion detection — they must be reconstructed by specialized neurons with large receptive fields. Each neuron collects a distinct subset of small field motion signals across the visual field and integrates them. When the arrangement of this subset resembles an optic-flow component, the integrating neuron may be regarded as a neural filter. In other words, it is specialized to extract that particular aspect of self-motion from the optic flow⁶. Each such neuron can be specified by the flow component that it responds to (either rotation or translation), and by the direction/orientation of the corresponding flow axis, or principal axis. It is from this type of neuron that Wylie *et al.* have been recording in the pigeon^{1,2}.

Freely flying animals experience an infinite number of different optic-flow fields, and it does not seem feasible to have specialized neural sensors for each of them. Instead, one can think of sets of neurons with orthogonal motion sensitivities. Intermediate directions of motion could then be interpolated, if need be. If this is so, how are the principal axes of the neurons that process optic flow orientated?

The vestibular system of the inner ear has a key function in sensing self-rotation. Its three semicircular canals (Fig. 1) are arranged orthogonally, in a very similar way in all vertebrates. One of their principal axes is vertical; the other two are on the horizontal plane and turned 45° and 135° from the forward direction. It has been shown that the principal axes of rotation-specific visual neurons coincide roughly with those of the canals^{2,7,8}. The preferred axes of eye rotation — given by the attachment of the extraocular muscles at the eyeball and the orbit — share the same orientation. This common frame of reference suggests that gaze-stabilizing eye movements in response to

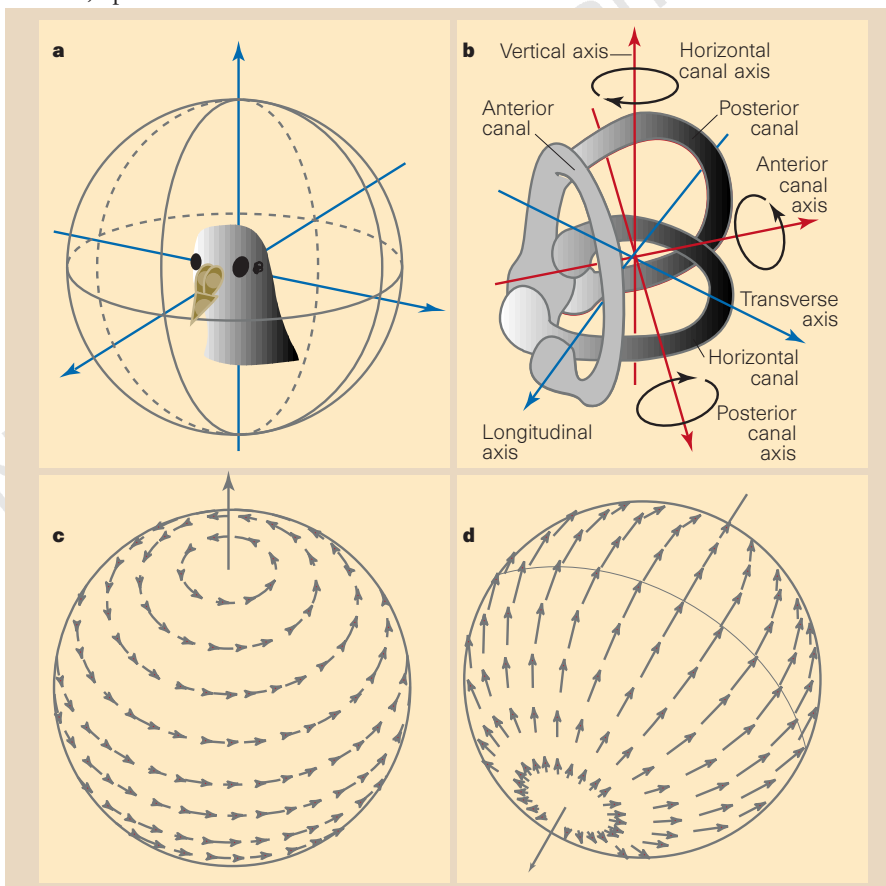


Figure 1 Analysis of optic flow in the pigeon brain, as studied by Wylie *et al.*¹. The nerve cells in the pigeon's brain that decipher the two types of optic flow have principal axes clustering around the same orientations as those of the semicircular canals. a, Head axes in the centre of the 'visual unit sphere', where all visible objects are assumed to be at the same, finite distance. b, Semicircular canals of the left inner ear, which sense rotations of the head and are orientated obliquely, relative to the head axes (blue), and roughly at right angles to one another. c, Rotatory optic flow, caused by a right-hand turn of the pigeon. d, Translatory optic flow, caused by forward progression of the pigeon in confined surroundings.

self-rotation can be made with a minimum of additional computation and, thus, very quickly.

Wylie *et al.*¹ have now shown that the pigeon also has translation-specific neurons to process optic flow. Interestingly, the orientation of their principal axes is similar to that of the rotation-specific cells. This is not a foregone conclusion — the translatable system could have been aligned with the body axes, or with the motor system that produces the translations. Instead, Wylie *et al.* show that both the rotatory and the translatable systems for sensing self-motion share a common frame of reference based on the arrangement of the vestibular system. This might simply be an accidental consequence of vertebrate evolution. It is more likely, however, that such a standard arrangement of different sensors has advantages for the processing of sensory information caused by

self-motion. We still need to demonstrate in detail what these advantages are. □

Roland Hengstenberg is at the Max-Planck-Institut für Biologische Kybernetik, Spemannstrasse 38, D-72076 Tübingen, Germany, and the Institute for Advanced Study, Wallotstrasse 19, D-14193 Berlin, Germany.

1. Wylie, D. R. W., Bischof, W. F. & Frost, B. J. *Nature* **392**, 278–282 (1998).
2. Wylie, D. R. & Frost, B. J. *J. Neurophysiol.* **70**, 2647–2659 (1993).
3. Koenderink, J. J. & van Doorn, A. J. *Biol. Cybern.* **56**, 247–254 (1987).
4. Borst, A. & Egelhaaf, M. in *Visual Motion and its Role in the Stabilization of Gaze* (eds Miles, F. A. & Wallman, J.) 3–27 (Elsevier, Amsterdam, 1993).
5. Amthor, F. & Grywacz, N. M. in *Visual Motion and its Role in the Stabilization of Gaze* (eds Miles, F. A. & Wallman, J.) 79–100 (Elsevier, Amsterdam, 1993).
6. Krapp, H. G. & Hengstenberg, R. *Nature* **384**, 463–466 (1996).
7. Simpson, J. I. & Graf, W. in *Adaptive Mechanisms in Gaze Control: Facts and Theories* (eds Berthoz, A. & Melvill-Jones, G.) 1–15 (Elsevier, Amsterdam, 1985).
8. Graf, W., Simpson, J. I. & Leonard, C. S. *J. Neurophysiol.* **60**, 2091–2121 (1988).

High-pressure physics

The double identity of ice X

José Teixeira

The position of hydrogen atoms along hydrogen bonds in ice is an almost unique source of information about some fundamental physical problems, such as chemical quantum effects, with implications in fields from material science to biophysics. The dynamics of hydrogen atoms in these bonds explains most of the unusual properties of liquid water (the strong isotope effects, for example, and the stability of the bonds formed between water molecules and substrates, such as amino-acid residues in proteins). We know that quantum effects are important¹; but on page 258 of this issue², Benoit *et al.* describe a surprising consequence of the quantum-mechanical behaviour of hydrogen atoms — the existence of two forms of symmetric ice.

The chemical formula of water tells us that two hydrogen atoms are attached to one

oxygen. Structural studies, on the other hand, show that each water molecule has four neighbours. This is exactly true for the common forms of ice, and approximately true for the liquid. Each molecule is linked to a neighbour by a hydrogen bond (a hydrogen is shared by two atoms — on one side by a covalent bond, on the other side by the

hydrogen bond). In ice, the position of the hydrogen is asymmetric, around 0.1 nm from one oxygen and 0.18 nm from the other. Along this O–O line, two positions are possible. In other words, the potential seen by the proton has two minima, and the proton is localized in one of them^{3,4}.

It was predicted 26 years ago that forcing the oxygen atoms closer together should make the two minima collapse into a single one, halfway between the two atoms^{5,6} (Fig. 1), yielding a new symmetric form of ice — ‘ice X’. But to achieve this, very high pressures are necessary, and it is only in the past few years that a few laboratories have been able to study samples at pressures near 100 GPa (a million atmospheres) in diamond anvil cells. The next step is determining the sample structure. X-ray or neutron scattering could directly determine the interatomic distances, but unfortunately hydrogen atoms are invisible to X-rays, and neutron scattering requires large samples. Consequently, the evidence for ice X has been indirect, coming from infrared^{7,8} and Raman⁹ spectroscopy: from the collapse of the three intramolecular vibrations into two lattice vibrational modes⁷, it can be inferred that ice X is formed.

The new computer simulation² of molecular dynamics, using path integral methods that treat the particles quantum mechanically, clearly establishes the existence of this new form of ice. Benoit *et al.* change the distance between two oxygen atoms continuously from 0.26 nm to 0.22 nm, which corresponds to an applied pressure of more than 140 GPa. At the beginning of this process the stable form is antiferroelectric ice VIII, and each proton is localized in one of the wells of the potential; by the end, the authors have ice X, and the proton is symmetrically placed between the oxygen atoms. At this point, the crystal is no longer molecular, but must be considered as an atomic crystal with two-thirds hydrogen and one-third oxygen atoms.

The surprise comes at intermediate pressures. There is, first, a transition to paraelectric ice VII, in which the energy barrier between the two wells is low enough to allow the proton to change site via tunnelling along the O–O line. At somewhat higher pressures it becomes centred midway along the O–O line due to zero-point motion (Fig. 2): quantum mechanics produces a symmetric form of ice earlier than expected, a different type of ice X.

The study of these two forms of ice X will be of crucial importance to liquid and solid water as well as to other hydrogen-bonded systems. In particular, the impact on biology may be important, namely in problems such as proton transfer through membranes. Also, in more fundamental areas, studies of the dynamics of hydrogen bonds, their characteristic lifetimes and breaking mechanisms in water and other hydrogen-bonded liq-

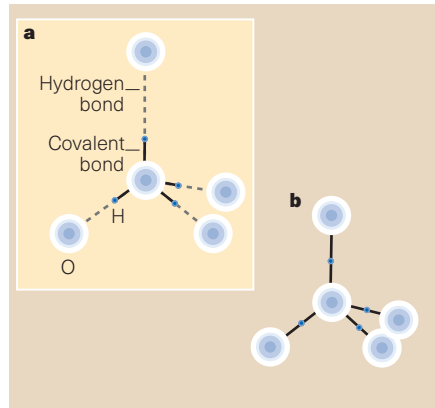


Figure 1 The structure of ice VIII (a) and of symmetric ice X (b). Applied pressure shortens the O–O distances in b, and the hydrogen atoms sit halfway between each neighbouring oxygen.

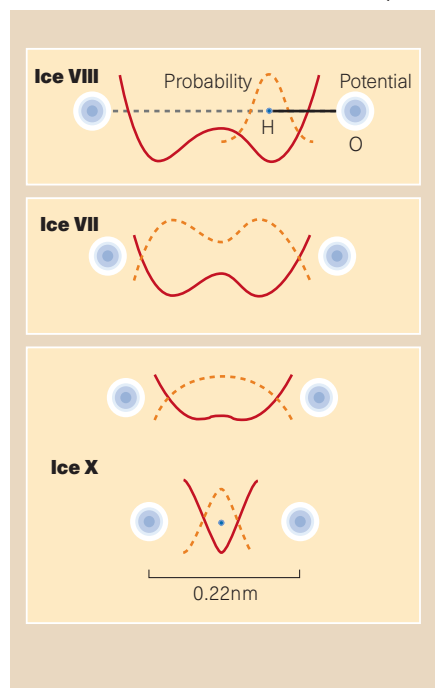


Figure 2 Proton density in ice under pressure. As the oxygen atoms are forced together, the potential energy changes from a double to a single well; but a new simulation shows that a form of ice with the proton midway between the oxygens occurs even before that happens.

uids¹⁰ will profit from the results obtained in the present simulation. □

José Teixeira is in the Laboratoire Léon Brillouin (CNRS/CEA), CE Saclay, 91191 Gif-sur-Yvette, France.

e-mail: teix@llb.saclay.cea.fr

1. Bratos, S., Diraison, M., Tarjus, G. & Leicknam, J.-C. *Phys. Rev. A* **43**, 5556–5565 (1992).
2. Benoit, M., Marx, D. & Parrinello, M. *Nature* **392**, 258–261 (1998).

3. Bernal, J. D. & Fowler, R. H. *J. Chem. Phys.* **1**, 515–548 (1933).
4. Pauling, L. *J. Am. Chem. Soc.* **57**, 2680–2684 (1935).
5. Ubbelohde, A. R. *Journal de Chimie Physique et de Physicochimie* **46**, 429–437 (1949).
6. Holzapfel, W. B. *J. Chem. Phys.* **56**, 712–715 (1972).
7. Aoki, K., Yamawaki, H., Sakashita, M. & Fujihisa, H. *Phys. Rev. B* **54**, 15673–15677 (1996).
8. Goncharov, A. F., Struzhkin, V. V., Somayazulu, M. S., Hemley, R. J. & Mao, H.-K. *Science* **273**, 218–220 (1996).
9. Pruzan, Ph., Chervin, J. C. & Canny, B. *J. Chem. Phys.* **99**, 9842–9846 (1993).
10. Luzar, A. & Chandler, D. *J. Chem. Phys.* **98**, 8160–8173 (1993).

Neuroscience

Right on in sign language

Eraldo Paulesu and Jacques Mehler

Language is a unique faculty of the human mind¹, and even the congenitally deaf can acquire it in the form of sign language². For more than a century — ever since the pioneering research of Paul Broca and Carl Wernicke, who first identified some of the key areas concerned — we have known that spoken language is represented in the brain's left hemisphere. But much remains mysterious about this lateralization of function. Is the left hemisphere concerned with language as a side-effect of the ability to process language-specific acoustic stimuli, or has grammar simply emerged there? Are all natural languages processed and represented in this part of the brain? Are all of the languages spoken by a polyglot processed by the same neural net? And are the same cortical areas used for the sign languages of the congenitally deaf and for the oral languages of the hearing?

A new study by Neville *et al.*, published last month in *Proceedings of the National Academy of Sciences*³, raises some provocative issues about the cerebral organization of

language in both deaf and hearing subjects. Neville and colleagues used functional magnetic resonance, a brain-imaging technique that can reveal the areas activated when subjects accomplish a task such as understanding sentences. They studied cerebral activity while subjects were processing sentences either in American Sign Language (ASL) or in written English.

Three groups of people were studied: (1) hearing monolingual speakers of English; (2) 'native' deaf signers (now properly called speakers) of ASL; and (3) hearing bilingual speakers of ASL and English. Subjects in group 3 were born to deaf parents and were exposed to sign language from birth; most had deaf siblings. The mastery of ASL in groups 2 and 3 was similar. Likewise, English proficiency was comparable in groups 1 and 3. The English proficiency of the deaf subjects in group 2 is said to have been only moderate; nevertheless, the three groups understood the English sentences presented during the experiment equally well.

During the scanning session, subjects watched two kinds of visual stimuli: first, videos of a deaf native ASL speaker producing sentences in ASL and, second, English sentences presented on a screen one word at a time. To identify the areas concerned with ASL processing or with reading English sentences, the activation due to the nonlinguistic properties of the stimuli was subtracted using two baseline tasks — looking at meaningless hand movements and meaningless strings of consonants.

As expected, the hearing volunteers with no knowledge of ASL showed no increase in brain activation when looking at ASL tapes. Moreover, Neville *et al.* observed an increase in activation in the left hemisphere when deaf or hearing subjects processed their native language (Fig. 1). These activated areas were located around the Sylvian fissure: that is, the inferior frontal gyrus (Broca's area), the superior temporal gyrus (including Wernicke's area) and also the left prefrontal cortex. More surprisingly, when processing sign language, native speakers of ASL (deaf and hearing) displayed a comparable increase of activation in equivalent areas of the right hemisphere. Finally, the deaf subjects showed a bilateral increase in activity while reading.

With the exception of the bilateral representation of sign language in native speakers of ASL, these findings are compatible with other observations involving hearing subjects^{4–6}. But although right-hemisphere activation associated with processing of oral and written language has been reported before^{4–6}, it was not as pronounced as activation of the left hemisphere. As far as we are aware, Neville and colleagues' report of almost symmetrical hemispheric language

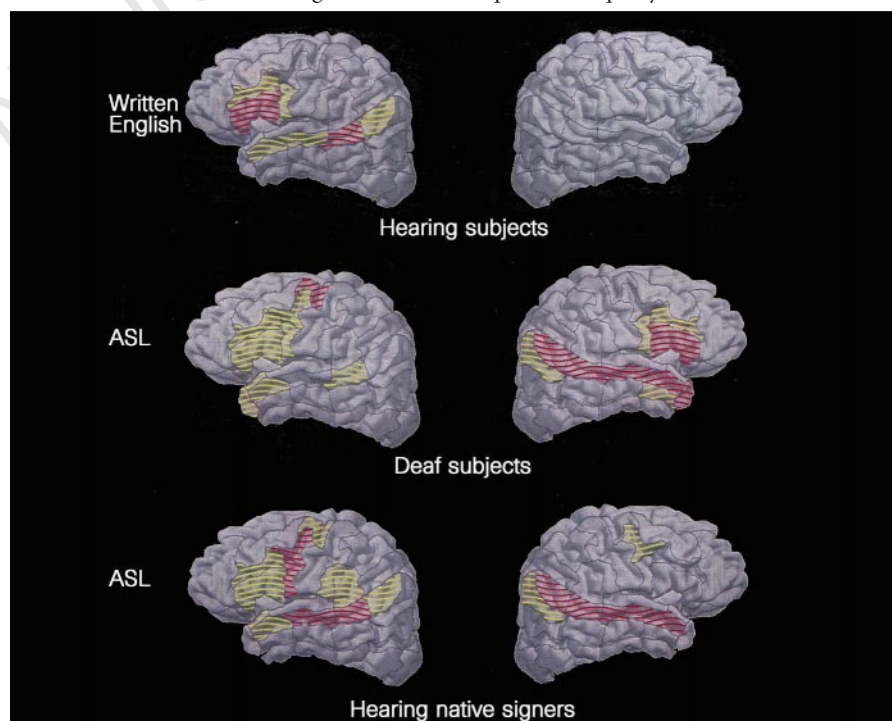


Figure 1 Summary of the findings of Neville *et al.*³ in imaging the areas of the brain activated when various subjects processed American Sign Language (ASL) or written English. The top row shows the areas activated when hearing monolingual speakers of English processed written English. The left hemisphere is much more active than the right. A similar pattern of activation was observed when a group of hearing native speakers (signers) of ASL processed written English. In comparison, a group of deaf ASL speakers showed less activation in the left hemisphere and also some activation in the right hemisphere (not shown). The second and third rows show the activation when sentences were processed in ASL: in both groups of native speakers of ASL, a bilateral pattern of activation was evident. As expected, the brain of monolingual non-ASL-speakers was not activated by ASL. Statistical significance of regional activation: red indicates $P < 0.0005$; yellow, $P < 0.005$; grey no activation. (Reproduced from Figs 1 and 2 of ref. 3.)

representation for sign language in hearing or deaf native speakers of ASL is unprecedented, and it is the most challenging aspect of their data.

For instance, the finding runs counter to observations with brain-damaged native speakers of ASL, in which injury to the left hemisphere (but not the right) often impairs their ability to use and understand ASL (aphasia)⁷. Given that Neville and colleagues' results imply that, in native ASL speakers, much of their right hemisphere is also activated by ASL, why have neuropsychologists failed to observe aphasia in such subjects when their right hemisphere has been damaged? This is an important question, but the uncertainty over the answer is nothing new — other results from imaging and from clinical neuropsychology have yielded partly inconsistent views about the anatomical foundations of language.

However, in this case there are some clues. For instance, Neville *et al.* show that the involvement of the right hemisphere is similar in deaf and hearing native speakers processing ASL sentences. So it seems that the involvement of the right hemisphere in processing ASL cannot result from a remapping of cortical areas because of deafness. Moreover, because the left hemisphere is dominant for oral languages, it is thought that it is the grammar of these languages that is mediated by the left hemisphere.

Could it be that the grammar for sign languages lies elsewhere? ASL uses hand and face movements, so one could imagine that the production and possibly the understanding of sign language require the representation of both sides of the body in the brain, resulting in bilateral activation. This suggestion conflicts, however, with the data of McGuire and colleagues⁸, who reported activation of the left prefrontal cortex and Broca's area in deaf speakers of sign language while they were covertly generating sentences in British sign language. They also described a similar pattern of activity when hearing subjects were silently generating English sentences⁹.

In explaining the discrepancy between the results of Neville and colleagues' functional imaging and findings from brain-damaged people, one cannot exclude the possibility that the right-hemisphere activation observed for ASL is due to a task-related variable that behaves differently for ASL and English. For instance, written English lacks features present in ASL and spoken languages, such as prosody (that is, rhythm and intonation). Deficits of prosody are seen following lesions to the right hemisphere¹⁰, and so reliance on this hemisphere may be different when reading English words and seeing ASL sentences.

Finally, the investigations of Neville *et al.* also raise questions about the representation of language not only in those people who are

bilingual for oral English and ASL, but also for other kinds of bilinguals. Among the authors' subjects were hearing users of ASL who were also highly proficient in English, whereas deaf users of ASL were only moderately proficient in English. The cortical activity of the deaf subjects for English sentences, especially in the left hemisphere, was less marked than in hearing ASL speakers. Age of language acquisition, degree of language proficiency and deafness are correlated in this comparison, so we cannot know which one is responsible for the different representations of English in the two groups of native ASL speakers.

Other investigations of the cortical representation of oral language in bilinguals suggest that proficiency in the second language is important in determining the overlap in the representation of the first and second languages^{11,12}. In people who have mastered two oral languages, there are only minute discrepancies between the cortical representations of the languages¹³. By contrast, the hearing native speakers of ASL and of English displayed considerable differences in the cortical representations of the two languages, even though their mastery of the languages was comparable to that of the best bilinguals in oral languages.

Could it be that ASL is implemented in such a unique sensory modality that its cortical representation is also unique, regardless of whether another language shares resources with it? To tackle this question, it might be necessary to explore further the behaviour of brain-damaged ASL speakers or momentarily to block the functioning of parts of the brain in order to understand exactly what their contribution is. A method such as trans-cranial magnetic stimulation¹⁴, which impedes the function of specified cortical areas, may be a useful tool to take this research further. □

Eraldo Paulesu is at the Scientific Institute H. San Raffaele, INB-CNR, Università di Milano, via Olgettina 60, 20132 Milan, Italy.

e-mail: epaulesu@mednuc.hsr.it

Jacques Mehler is at the Laboratoire de Sciences Cognitives et Psycholinguistique, CNRS et EHESS, 54 Boulevard Raspail, 75006 Paris, France.

e-mail: mehler@lscp.ehess.fr

- Chomsky, N. *Reflections on Language* (Pantheon, New York, 1975).
- Bellugi, U., Poizner, H. & Klima, E. *Trends Neurosci.* **12**, 380–388 (1989).
- Neville, H. J. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 922–929 (1998).
- Mazoyer, B. M. *et al.* *J. Cogn. Neurosci.* **5**, 467–479 (1993).
- Perani, D. *et al.* *NeuroReport* **7**, 2439–2444 (1996).
- Bottini, G. *et al.* *Brain* **117**, 1231–1253 (1994).
- Hickok, G., Bellugi, U. & Klima, E. *Nature* **381**, 699–702 (1996).
- McGuire, P. *et al.* *NeuroReport* **8**, 695–698 (1997).
- McGuire, P., Silbersweig, D. & Frith, C. *Brain* **119**, 907–917 (1996).
- Weintraub, S., Mesulam, M. & Kramer, L. *Arch. Neurol.* **38**, 742–744 (1981).
- Dehaene, S. *et al.* *NeuroReport* **8**, 3809–3815 (1997).
- Perani, D. *et al.* *J. Neurosci.* **23**, S2228 (1997).
- Kim, K., Relkin, N., Lee, K.-M. & Hirsch, J. *Nature* **388**, 171–174 (1997).
- Cohen, L. G. *et al.* *Nature* **389**, 180–183 (1997).

Daedalus

Free captive fish

The fishing industry is approaching a crisis. Once-rich fishing grounds such as the Grand Banks off Newfoundland are denuded; important species such as cod are becoming endangered; national fishing fleets argue ever more bitterly about fishing rights. And yet the obvious answer, fish farming, has problems of its own. For water is poor in food and oxygen, so fish do best in high dilution. Yet a commercial farm must pack its fish as densely as possible: depleting the water of oxygen, polluting it with waste and faecal matter, and maybe spreading disease. Daedalus now has a middle way.

Many fish farms 'wall off' a region of natural water with a net, and breed the fish behind it. Daedalus is taking this idea to extremes. His plan is to raise a shoal of fish to young adulthood in a farm, and then release them within a large net in the open ocean. He is designing a floating circular boom, suspended from which a huge closed net encloses a volume of sea like a vast tea-strainer. The mesh size is chosen so that the fish within the net cannot escape, and their predators (which must be bigger) cannot get at them; neither are they likely to be seriously invaded by unwanted species. At the same time, their food or prey (which must be smaller) can freely enter, and excreta can freely drift out. And, of course, clean bulk sea water will swirl through the net as if it were not there.

The final step is to set the whole thing adrift in the open sea. The fish will not even realize that they are trapped; they will behave like a free-living shoal. But radar reflectors and transponders on the circular boom will tell the owners where it is at any moment. When the fish have had time to grow big enough, a boat can go out to harvest them — either hauling in the net, or towing the whole thing back to port.

An unpowered 'fish corral' will drift freely with the current, and will not sustain a healthy flow of water through it — unless the fish themselves propel it, as they may. If not, its radar reflectors may have to be formed as sails, to drive it through the water. Remote control of those sails could even steer it about, though this erodes the simplicity of the scheme. Many details remain to be worked out; the optimum size of the fish corrals, the best locations for them, and ways of avoiding collisions with ships or hijacking by pirates. But when perfected, Daedalus's fish corrals should avert oceanic ecological disaster while providing a sustainable, predictable supply of healthy and nutritious fish.

David Jones

De Vries's varieties

Scientists have always had an irresistible urge to list and classify. But the work of the artist Herman de Vries reminds us that nature's variety will always defy our attempts at imposing a fixed order.

Martin Kemp

The history of natural history revolves around acts of naming and classification. This has most obviously been the case in the Linnaean and Darwinian eras, in which the concept of the species assumed such centrality. But it is equally true of those eras in which the name of something served to inscribe it in the book of nature and declared its symbolic significance within God's treasury of wonders.

Naming and taxonomy are tools, functionally apposite according to the demands of the users. The systems accordingly vary across different cultures, by place and time.

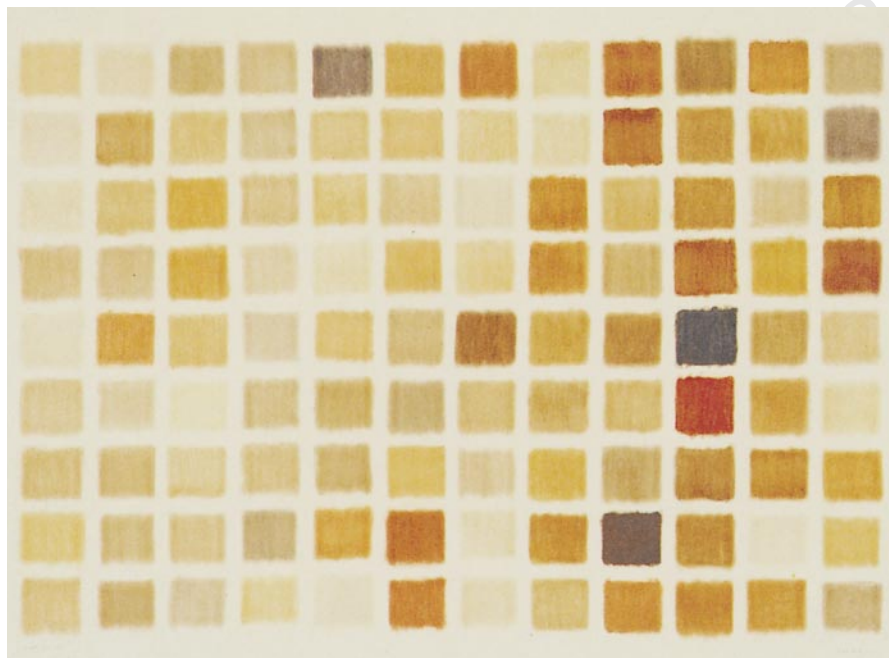
Herman de Vries, the Dutch artist, philosopher and poet, is a collector of natural things — things of a kind. But his 'kinds' stand in a complex and often challenging relationship to standard taxonomic categories, such as those in the herbaria and aboreta in which he has worked.

Since 1970 he has been exhibiting collections — of shells, flowers, leaves, branches, herbs and varieties of earth — which can sometimes be packaged intellectually in the standard fashion but which often resist the process of separation that denomination entails. So his montages of earths alert us to a multitudinous variety which far outstrips our leaden vocabulary.

He has a collection of more than 2,400 samples of earth, including no less than 300 from Gröningen in Holland and 350 from the volcanic island of Gomera in the Canaries. They are laid out according to criteria of similarity and difference, but not within rigid categories. The visual differences in the 'element' (in the ancient sense) that is the basis of so much life speak by implication of the manifold variations in the lives of plants and of the wonderful attuning of their physiologies to minute variations in the compositions of soils.

De Vries's art involves the quiet contemplation of nuances. But it is also quietly subversive. When he exhibits collections of herbal substances, rich in odour, he plays to sensory perceptions even more resistant to exact naming than objects of sight. When he records the sounds of the world he alerts us to a natural music which does not lend itself to conventional notation.

When he writes without capital letters, as he has done for 20 years, he intends to signal



De Vries's *from earth; nepal and india* (earth on paper), 1992.



De Vries's *forms from the botanic garden* (leaves on paper), 1991–92.

“a kind of anti-hierarchic expression. it's the same in nature; every part has its own function, so why should a tree be more important than a diatom?... language is ‘you and me’, ‘we and them’, ‘here and there’, whereas in effect, it is all part of the same, it's one. language... gives us a grip on reality and great social power, but we also pay for it in the loss of unity.”

His book of nature has no rigidly pre-

scribed lists of flora and fauna. He invites us to sense anew, so that we can sense affinities across boundaries. He also alerts us to pages torn out — as in his book *in memory of scottish forests* — forests still named on some maps but whose absence testifies to the perils of collective insensitivity.

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

Carbon nano-cages created as cubes

We have produced a new form of graphitic nano-cages, rectangular parallelepipeds or cubes, by arc evaporation of carbon with the alkaline-earth metals calcium or strontium. The cubes contain 5 to 20 layers of multiwalled graphitic carbon and have edges ranging in length from 20 to 100 nanometres. Their shape can be controlled by the type of metal catalyst selected.

The method used to obtain these cubes was the same as that for synthesizing carbon nanocapsules filled with rare-earth^{1,2} and iron-group metals³: that is, evaporating a metal-loaded graphite rod (anode) in a helium atmosphere by a direct-current arc discharge. The metal-loaded anode was prepared by packing a hole (measuring 30 mm deep by 3.2 mm in diameter) drilled in a graphite rod (50 mm long by 6 mm diameter) with small pieces of calcium (purity 99%) or strontium (purity 99%). The graphite used for the rod was of 99.998% purity, as was that for the 13-mm cathode. The helium (purity 99.999%) introduced into the arc chamber had a pressure of typically 100 and 600 torr. Discharge current and voltage were 70 amp and about 25 V respectively.

After arc evaporations using both calcium and strontium, we found, in the soot deposited on the cathode surfaces, abundant cubic cages of graphitic carbon (Fig. 1a). The cages range in size from 20 to 100 nm and consist of multiwalled graphitic carbon with a spacing of 0.34 nm (Fig. 1b). Adjacent graphitic layers were out of register, or turbostratic. At some of the cubes' corners, extrusion of graphitic layers can be seen, while breakages or smooth folding occurs at other corners (arrows in Fig. 1b). Most of the hollow cubes contain smaller cubes, which are also hollow.

Some other shapes of nano-cage — fullerenes⁴ and nanotube tips⁵ — are closed by the introduction of 12 pentagons into the hexagon network. But in the case of a

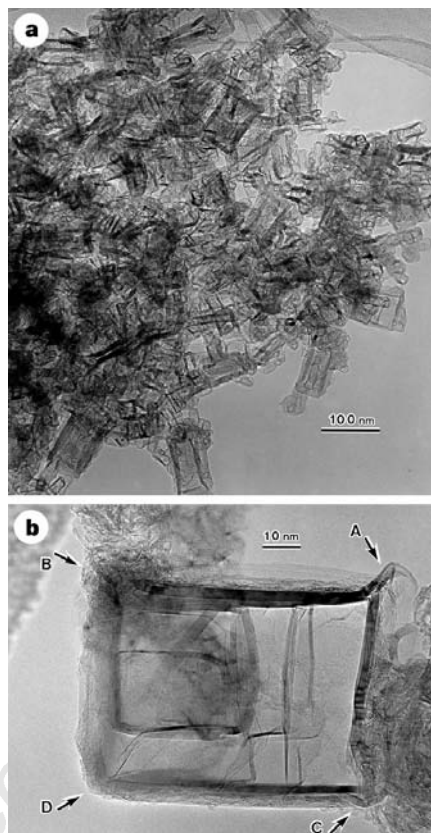


Figure 1 Transmission electron microscope images of hollow rectangular parallelepiped graphitic cages formed by the arc evaporation of a carbon/calcium composite. **a**, High-density aggregation of hollow rectangular parallelepipeds (low magnification). **b**, Aggregation of multiwalled graphitic carbon with a spacing of 0.34 nm (high resolution). At some corners, extrusion (arrows A, B) and breakage (arrow C) of graphitic layers can be seen, and sometimes graphitic layers are folded continuously (arrow D).

rectangular parallelepiped cage with eight vertices, pentagons or even four-membered rings cannot effect closure. Therefore, the rectangular corners suffer from geometrical frustrations, which may be relaxed by either

extruding or discontinuing (breakage) graphitic layers at the corners.

The cubic shape is probably produced by folding graphitic sheets into a rectangular form, a process catalysed by alkaline-earth metals. But questions remain over the detailed structures at the corners and the role of metallic atoms.

Although the hollow cubic cages predominated, we also found square cages filled with foreign materials. Most of the filling crystallites were carbides (CaC_2 and SrC_2), but sometimes metallic strontium was encapsulated. Although these carbides and metal are hygroscopic, the crystallites nesting in the graphite cages did not degrade even when they were exposed to air. The closed graphitic cages effectively shield the inner materials from moisture, in a similar way to the carbon nanocapsules with polyhedral^{1,2} and spherical shapes³ that have been reported previously.

Graphitic carbon is found in forms and shapes as various as single-walled^{6,7} and multiwalled tubules⁸, cones⁹, polyhedra^{1,2} and spheres^{3,10}. The topology of graphite networks governs the physical properties of nanometre-sized graphitic materials, so our success in producing a new form of graphitic cages should provide further opportunities for exploration of their exotic and unique properties.

Yahachi Saito, Takehisa Matsumoto

Department of Electrical and Electronic Engineering, Mie University, Tsu 514, Japan
e-mail: saito@is.elec.mie-u.ac.jp

- Tomita, M., Saito, Y. & Hayashi, T. *Jpn. J. Appl. Phys.* **32**, L280–L282 (1993).
- Ruoff, R. S. *et al. Science* **259**, 346–348 (1993).
- Saito, Y. *et al. J. Phys. Chem. Solids* **54**, 1849–1860 (1993).
- Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162–163 (1985).
- Iijima, S., Ichihashi, T. & Ando, Y. *Nature* **356**, 776–778 (1992).
- Iijima, S. & Ichihashi, T. *Nature* **363**, 603–605 (1993).
- Bethune, D. S. *et al. Nature* **363**, 605–607 (1993).
- Iijima, S. *Nature* **354**, 56–58 (1991).
- Sattler, K. *Carbon* **33**, 915–920 (1995).
- Ugarte, D. *Nature* **359**, 707–709 (1992).

Tight knot values deviate from linear relations

Applications of knots to the study of polymers have emphasized geometric measures on curves such as 'energy'^{1–4} and 'rope length'^{5–7}, which, when minimized over different configurations of a knot, give computable knot invariants related to physical quantities⁸. In DNA knots, electrophoretic mobility appears to be correlated with the average crossing number of rope-length-minimizing configurations⁹, and a roughly

linear empirical relation has been observed between the crossing number and rope length¹⁰. Here we show that a linear relation cannot hold in general, and we construct infinite families of knots whose rope length grows as the $3/4$ power of the crossing number¹¹. It can be shown that no smaller power is possible^{12–14}.

One measure of geometric complexity for a space curve γ is $A(\gamma)$, the average number of crossings in planar projections of γ . Another scale-invariant measure is rope length $L(\gamma)$, the quotient of arclength by thickness (the diameter of the largest uniform tube centred on γ).

If we fix a knot type K (or a link type K : our methods work for curves with one or many components), the infima (greatest lower bounds) of A and L over γ in K are the crossing number $C(K)$ and rope length $L(K)$ of K . A curve γ achieving the infimum $L(K)$ is called 'tight'. Experiments⁹ with knotted DNA and simulations¹⁰ of thermal averages of such knots led us to seek mathematical relations between the invariants C and L , and between A and L for tight curves.

For any fixed power $3/4 \leq p \leq 1$ we have constructed infinite families of knots and of links that have $L \sim C^p$ — the ratio L/C^p is

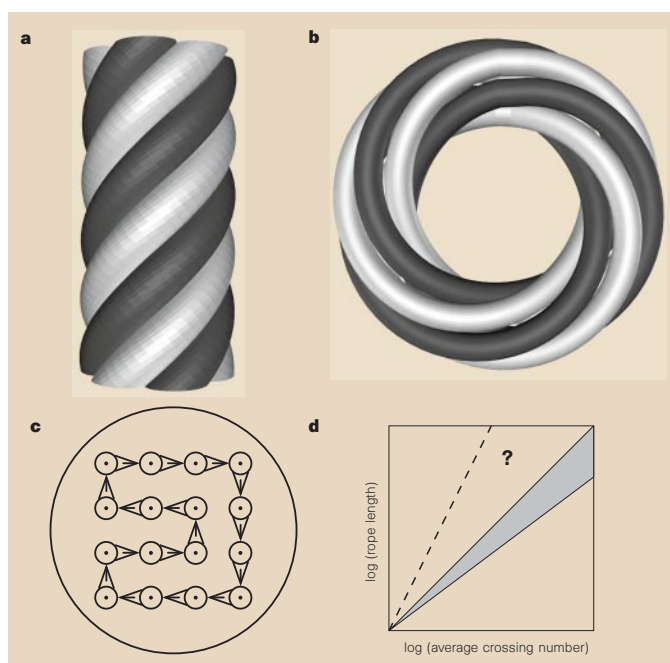


Figure 1 Construction for links with rope length L growing as the $3/4$ power of crossing number C . **a**, This 'braid' has seven helical strands, one hidden in the centre. **b**, The braid can be bent into this Hopf link. **c**, The construction can be modified by following this circuit to produce a torus knot instead. **d**, The general relation between crossings and rope length when viewed at large scale: for any family of links, the growth rate must lie between powers $3/4$ and 2 , and all powers below 1 are realized.

bounded above and below across each family. No $p < 3/4$ can work, because $C \leq A$ and there is a universal lower bound¹² for $L/A^{3/4}$. Furthermore, rope-length minimizers in a family with $L \sim C^{3/4}$ must satisfy $C \sim A$; thus these tight curves have $L \sim A^{3/4}$.

Any knot or link arises from a 'braid'; a collection of N ascending arcs in a cylinder Z , joining N points on the bottom of Z to the same N points on the top. Z is bent into a solid torus S , the top and bottom disks are identified, and the arcs are joined to give a link in S . For a braid with bounds on its slope, curvature and horizontal strand separation, this bending fixes rope length within a uniform factor (depending on the shape of Z and S).

For a family of N -component Hopf links with $L \sim C^{3/4}$, we took a cylinder Z of height h and radius r . N points are distributed on the bottom disk of Z , and rotated one full turn as they ascend. The resulting braid (Fig. 1a) has N helical strands of bounded slope and curvature, provided $h \approx r$, to separate strands, we needed $r \sim N^{1/2}$. Thus the braid (and the corresponding link, Fig. 1b) has rope length $L \sim hN \sim N^{3/2}$. As each component links every other component exactly once, $C \geq N(N-1)$; the standard projection has $N(N-1)$ crossings, so $C = N(N-1) \sim N^2$.

For $(N, N-1)$ -torus knots in S with $L \sim C^{3/4}$, we repeated this construction in the lower half of Z . To define the rest of the braid, we chose a circuit joining the N points (Fig. 1c) and slid each point along this circuit to the next while ascending through the upper half of Z . The horizontal distance a point travels is comparable to, at most, r . So, as before, we needed $h \sim r \sim N^{1/2}$ and rope length $L \sim N^{3/2}$. The minimum crossing number of this knot¹⁵ is $C = N(N-2) \sim N^2$. (Lattice knots with the same growth of crossing number have been

found independently¹⁶.) Many further families of links can be thus constructed¹¹.

If we plot C versus L on a log-log scale (Fig. 1d), then links in any family with $L \sim C^p$ approach the ray of slope p . We have seen examples with $p = 3/4$ (sublinear growth). To get $p = 1$ (linear growth), consider simply linked chains of N components, for which $L = 2\pi N + 2(N-2)$, whereas $C = 2(N-1)$ (ref. 11). Combining these examples yields families of links with any $3/4 \leq p \leq 1$.

It is unknown whether any family has superlinear growth. But if links with split, unknotted components are excluded, an argument based on embedding planar projections¹¹ shows that no family has growth $p > 2$.

Jason Cantarella

Department of Mathematics,
University of Pennsylvania,
Philadelphia, Pennsylvania 19104, USA

Robert B. Kusner

School of Mathematics, Institute for Advanced Study,
Princeton, New Jersey 08540, USA
(Permanent address: Department of Mathematics,
University of Massachusetts, Amherst,
Massachusetts 01003, USA)

John M. Sullivan

Department of Mathematics,
University of Illinois, Urbana, Illinois 61801, USA
e-mail: jms@math.uiuc.edu

- O'Hara, J. *Topology* **30**, 241–247 (1991).
- Freedman, M. H., He, Z.-X., & Wang, Z. *Ann. Math.* **139**, 1–50 (1994).
- Kim, D. & Kusner, R. *Exp. Math.* **2**, 1–9 (1993).
- Kusner, R. B. & Sullivan, J. M. in *Geometric Topology* (ed. Kazez, W. H.), 570–604 (Am. Math. Soc./Int. Press, Providence, 1997).
- Kusner, R. B. & Sullivan, J. M. in *Topology and Geometry in Polymer Science* (eds Whittington, S., Sumners, D. & Lodge, T.), 67–78 (Springer, New York, 1998).
- Litherland, R. A., Simon, J., Durumeric, O. & Rawdon, E. *Topol. Appl.* (in the press).
- Nabutovsky, A. *Comm. Pure Appl. Math.* **48**, 381–428 (1995).
- Moffatt, H. K. *Nature* **384**, 114 (1996).
- Stasiak, A., Katritch, V., Bednar, J., Michoud, D. & Dubochet, J. *Nature* **384**, 122 (1996).
- Katritch, V. *et al. Nature* **384**, 142–145 (1996).
- Cantarella, J., Kusner, R. B. & Sullivan, J. M. (Preprint, Dept of Mathematics, Univ. of Illinois, 1997).
- Cantarella, J., DeTurck, D. & Gluck, H. (Preprint, Dept of Mathematics, Univ. of Pennsylvania, 1997).
- Freedman, M. H. & He, Z.-X. *Ann. Math.* **134**, 189–229 (1991).
- Buck, G. & Simon, J. *Topol. Appl.* (in the press).
- Murasagi, K. *Trans. Am. Math. Soc.* **237**–260 (1991).
- Diao, Y. & Ernst, C. *Topol. Appl.* (in the press).

Four-thirds power law for knots and links

Physical knot theory has recently been applied to polymer dynamics, and specifically to gel electrophoresis of DNA^{1,2}. Knot energies^{3–6} measure the complexity of a knot conformation; minimum energy conformations are considered canonical or 'ideal' conformations. The rope length of a knot is one such measure of energy⁶, and an approximately linear relationship between rope length and the average crossing number for minimum rope-length conformations of simple knots has been reported⁷. Here I show that a linear relationship cannot hold in general: the rope length required to tie an N -crossing knot or link varies at least between $\sim N^{3/4}$ and $\sim N$.

Consider four measures of knot conformation complexity. Imagine a solid disc of radius R centred at x and normal to K at each point x along the parametrized smooth knot K . $R(K)$ is the largest R so that the disks are disjoint. The rope length is $L(K) = [\text{arclength}(K)]/[R(K)]$. The crossing number $C(K)$ of the knot-type K is the necessary number of crossings in the planar diagram of K .

For points x, y on K , let ρ denote $|x - y|$, r the vector $(x - y)/|x - y|$, and dx a line element of K . The average crossing number is given by

$$A(K) = (1/4\pi) \iint_{K \times K} (|dx, dy, r|) / \rho^2$$

where the integrand numerator is the positive triple scalar product of the three vectors. This integral gives the average number of crossings of K , when viewed from an arbitrary perspective. The symmetric energy³ is given by

$$S(K) = \iint_{K \times K} (|dx \times r| |dy \times r|) / \rho^2$$

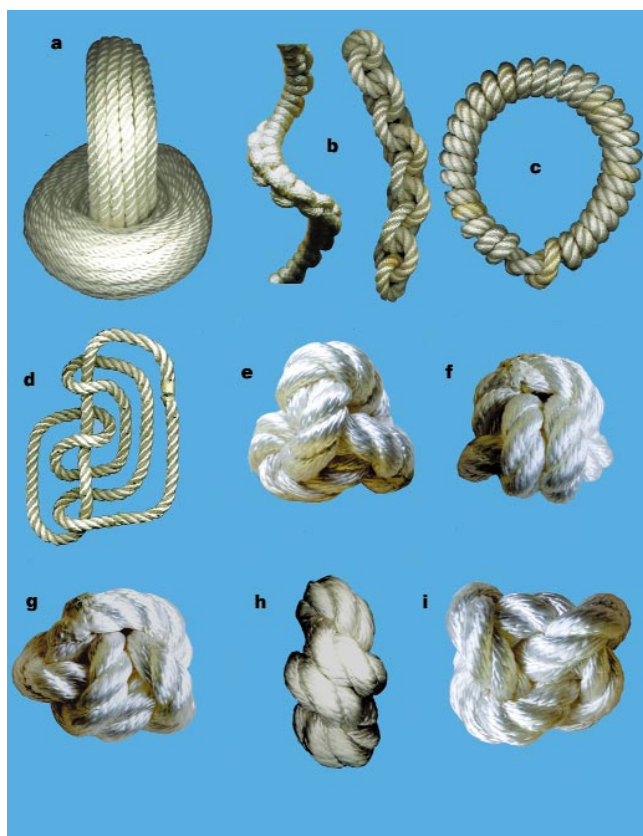
If K is the centre curve of a radiating tube, T , with relatively small radius, which is relatively far from self-intersection, then $S(K)$ measures the self-radiation of the tube.

A natural $4/3$ power law relates rope length to measures based on inverse-square laws (including the crossing and writhe numbers, and $S(K)$), and a linear bound relates $S(K)$ to A . We can show^{4,17} that

$$11L(K)^{4/3} \geq S(K) \geq 4\pi A(K) \geq C(K)$$

Imagine a very long piece of rope packed as tightly as possible in a roughly spherical

Figure 1 Knot conformations. **a**, Packed Hopf tori. **b**, 'Linear' conformations. Left, a product of trefoils; right, a thick chain with a linear relationship between crossing number and rope length. The chain also seems to be a continuous family of minima for rope length, in which case minima are not isolated in the link class. **c**, 'Linear' conformation of a twist knot – apparent minimum. **d**, An N -crossing knot fits in a square of side order N . **e, f**, Minima for figure-eight and square knot respectively. No particular accuracy is claimed – these knots were tied before the calculation of the computer data⁷, and both the conformation and the values for rope length match almost exactly, as did several other knots. **g**, Minimum for the 'granny' knot, differing in shape from the minimum found by computer⁸, and having a different symmetry. **h**, Another view. We estimate this to be the true global minimum. **i**, Some support for the accuracy of the rope calculations is given by this conformation of the five-crossing torus knot, which by rope seemed to be a lower minimum than the conformation reported in ref. 7. Further computation confirmed this.



shape, scaled so that $R(K)=1$, with total length L . The inverse-square 'energy' ($S(K)$, A , writhe, and so on) can be estimated by assuming the 'mass' of the knot is concentrated at points p on the integer lattice. Concentric shells of unit thickness about each p each contribute the same amount, so the contribution for p is that constant multiplied by the number of shells, which is of the order of $L^{1/3}$. Multiplying by the number of points, L , gives $L^{4/3}$. The proof that $S(K)$ linearly bounds A is simple vector geometry.

The $4/3$ exponent is sharp. Consider the Hopf link of two tori in its natural geometrical position. Fill each torus with N loops parallel to the centre curve, each loop a strand of radius 1 (Fig. 1a). Then with any tight packing of the loops, the minor radii of the tori is of the order of $\sqrt{N}$. The conformation fits inside a sphere of radius $4\sqrt{N}$, so the total rope length is about $N^{3/2}$. Each loop is linked with N loops in the perpendicular torus, so the crossing number is about N^2 . Therefore the rope length is of the order of $C(K)^{3/4}$. Because $11L(K)^{4/3} \geq 4\pi A(K) \geq C(K)$, this example has A in the order of $L(K)^{4/3}$.

The minimum rope length for a knot is bounded by $3C(K)^2$. This can be seen by arranging the knot so that the crossings are evenly spaced along a line (Fig. 1d). For the simpler knot types, $L(K)$, $S(K)$ and A in minimized conformations all 'appear' to be

linearly related⁷. An explanation is that the simpler conformations are 'planar': from most perspectives a unit arc of the knot crosses only a few other unit arcs.

As complexity increases, there are many families of knots and links with three-dimensional growth, exhibiting the $4/3$ power law. Families with single-dimensional growth (Fig. 1b,c) have a linear relationship among the measures. With planar growth, we expect A to be linear with $C(K)$ and $S(K)$ to be of the order of $L(K)\log L(K)$.

We propose that the rope length required (Fig. 1e–i) to tie an N -crossing knot or link varies only between $k_1 N^{5/4}$ and $k_2 N$. Other investigators have also recently observed the $4/3$ law in knots on the cubic lattice⁹ and in vector fields¹⁰.

A good knot energy has only a finite number of knot types realized below any given energy level. Our theorem gives us this property for $L(K)$ and $S(K)$, proving that there is a finite number of knots that can be tied with a finite length of mathematical rope.

Gregory Buck

Department of Mathematics, Saint Anselm College,
Manchester New Hampshire 03102, USA
email: gbuck@anselm.edu

1. Simon, J. in *Mathematical Approach to Biomolecular Structure and Dynamics* (Proc. of 1994 IMA Summer Program on Molecular Biology; eds Mesirov, J. P., Schulten, K. & Sumners,

D.W.) 39–58 (IMA 82, Springer, New York, 1996).

2. Stasiak, A., Katritch, V., Bednar, J., Michoud, D. & Dubochet, J. *Nature* **384**, 122 (1996).
3. Buck, G. & Simon, J. *Topol. Appl.* (in the press).
4. O'Hara, J. *Topology* **30**, 241–247 (1991).
5. Moffatt, H. K. *Nature* **347**, 367–369 (1990).
6. Buck, G. & Orloff, J. *Topol. Appl.* **61**, 205–214 (1995).
7. Katritch et al. *Nature* **384**, 142–145 (1996).
8. Buck, G. (Preprint no. 2, Math. Dept, St Anselm College, 1993).
9. Diao, Y. & Ernst, K. *Topol. Appl.* (in the press).
10. Cantarella, J., DeTurck, D. & Gluck, H. (Preprint, Math. Dept., Univ. Pennsylvania, 1997).
11. Katritch, V., Olson, W. K., Pieranski, P., Dubochet, J. & Stasiak, A. *Nature* **388**, 148–151 (1997).

Heartbeat synchronized with ventilation

It is widely accepted that cardiac and respiratory rhythms in humans are unsynchronized¹. However, a newly developed data analysis technique allows any interaction that does occur in even weakly coupled complex systems to be observed. Using this technique, we found long periods of hidden cardiorespiratory synchronization, lasting up to 20 minutes, during spontaneous breathing at rest.

Synchronization is a universal phenomenon that occurs when two or more nonlinear oscillators are coupled. It is observed in many fields of science and is widely applied in engineering. The case of synchronisation in periodic, or even noisy, oscillators is well understood^{2–4}. The notion of synchronization has often been used to analyse the interaction between physiological (sub)systems¹, but these studies have been restricted to almost periodic rhythms. No approach has been suggested to probe the weak interactions between such irregular and non-stationary oscillators as the human heart and respiratory system.

These two physiological systems are known to be coupled by several mechanisms, but apart from respiratory modulation of heart rate, first observed in 1847 and known as 'respiratory sinus arrhythmia' (RSA)^{5–7}, no other effects have been reported. Moreover, in spite of some early communications⁸, it has been concluded that "there is comparatively weak coupling between respiration and the cardiac rhythm, and the resulting rhythms are generally not phase locked"¹.

We used the concept of phase synchronization of chaotic oscillators^{9,10} to develop a technique to analyse irregular non-stationary bivariate data. We analysed data obtained in non-invasive examinations of eight healthy volunteers (14–17-year-old, high-performance swimmers; four of them male and four female). While subjects lay at rest, electrocardiograms (ECGs) were recorded while respiratory

Table 1 **Subjects, ordered by the amplitude of respiratory sinus arrhythmia (RSA) determined as the averaged difference between the longest and shortest R–R interval within every respiratory cycle***

Code	Sex	Age	R–R (ms)		Resp. cycle (ms)		RSA (ms)		Synchronization
			Median	Δ Quart.	Median	Δ Quart.	Median	Δ Quart.	
A	m	16.1	1,104	28	3,110	390	15	40	3:1 (1,000 s), 5:2 (300 s), 8:3 (20 s)
B	m	14.6	1,018	95	3,210	610	31	38	3:1 (several spells of 40 s)
C	m	13.9	975	110	3,230	850	46	57	3:1 (20 s), 7:2 (20 s), 4:1 (20 s)
D	f	15.2	1,157	157	2,930	780	56	57	5:2 and 3:1 (several spells of 30 s)
E	m	16.9	1,026	89	3,650	620	67	47	7:2 (60 s), 3:1 and 4:1 (20s)
F	f	15.0	1,024	143	2,960	700	74	75	11:4 (20 s)
G	f	15.9	733	70	5,615	1,550	83	70	No synchronization detectable
H	f	16.3	1,256	197	4,260	2,100	264	296	No synchronization detectable

*If an R–R interval spans two adjacent cycles, it is considered to belong to that one which contains more than 50% of the interval. For R–R intervals, respiratory cycles and the RSA, the medians of respective distributions and differences between the first and third quartile (Δ Quart.) are given. We observe that cardiorespiratory synchronization tends to become weak with increasing RSA.

flow was simultaneously measured with a thermistor at the nose. Both signals were digitized with a 1,000-hertz sampling rate and 12-bit resolution. Each record lasted 30 minutes.

The resulting time series were irregular, strongly non-stationary and noisy. These characteristics made it inappropriate, in analysing the mutual dependencies involved, to use even sliding versions of traditional spectral and correlations techniques, or direct computation of instantaneous phase differences^{9,10}. So instead of these techniques, we used a new kind of data presentation which we call a cardiorespiratory synchrogram (CRS), to detect different synchronous states and transitions between them.

We first used the Hilbert transform⁹ to calculate the instantaneous phase $\phi_i(t)$ of the respiratory signal. $\phi_i(t)$ is defined on the real line (not restricted to the $[0, 2\pi]$ interval). Next, we regarded the respiratory

phase stroboscopically at times t_k , where the R-peak in the k th heartbeat occurs and hence the phase of the heart rhythm increases by 2π . In the simplest case of n :1 synchronization, there are n heartbeats in each respiratory cycle; these beats appear at n certain values of respiratory phases, which are constant over all cycles.

Plotting these relative phases ψ as a function of time shows n horizontal stripes. In the general case of n : m synchronization, such a structure appears if we relate the phases of the heart beats to the beginning of m adjacent respiratory cycles, $\psi(t_k) = (\phi_i(t_k) \bmod 2\pi m) / 2\pi$; we have n horizontal stripes within m respiratory cycles.

This technique allows us to distinguish between different periods of synchronization, even for noisy and non-stationary data. For example, we observe 5:2 locking between the respiratory frequency ω_i and the heart rate ω_h ($5\omega_i \approx 2\omega_h$) during a

period of about 300 seconds, then after a transition period, a regime of 3:1 phase locking is found for about 20 minutes (Fig. 1). These two kinds of locking can be recognized using histograms (Fig. 1c) and the autocorrelation function of phases (Fig. 1d).

Our analysis showed cardiorespiratory synchronization of several locking ratios occurring in six out of eight subjects (Table 1). Subjects with the strongest synchronization had no remarkable RSA, whereas both subjects with the highest RSA exhibited no synchronization.

We conclude that phase locking of respiratory and the cardiac rhythms, and respiratory modulation of heart rate, are two competing aspects of cardiorespiratory interaction. From a physical viewpoint, synchronization and modulation are different phenomena and are related to different coupling. RSA generation is associated mainly with the baroreflex feedback loop and its respiratory phase-dependent information processing⁷.

Perhaps cardiorespiratory synchronization is an expression of another kind of cardiorespiratory interaction, such as a central coupling between cardiovascular and respiratory neuronal activities.

Carsten Schäfer, Michael G. Rosenblum, Jürgen Kurths

Department of Physics, University of Potsdam,
Am Neuen Palais 10, PF 601553,
14415 Potsdam, Germany
e-mail: carsten@agnld.uni-potsdam.de

Hans-Henning Abel

Department of Cardioanaesthesiology,
Städt Klinikum Braunschweig,
Salzdahlumer Strasse 90,
38126 Braunschweig, Germany

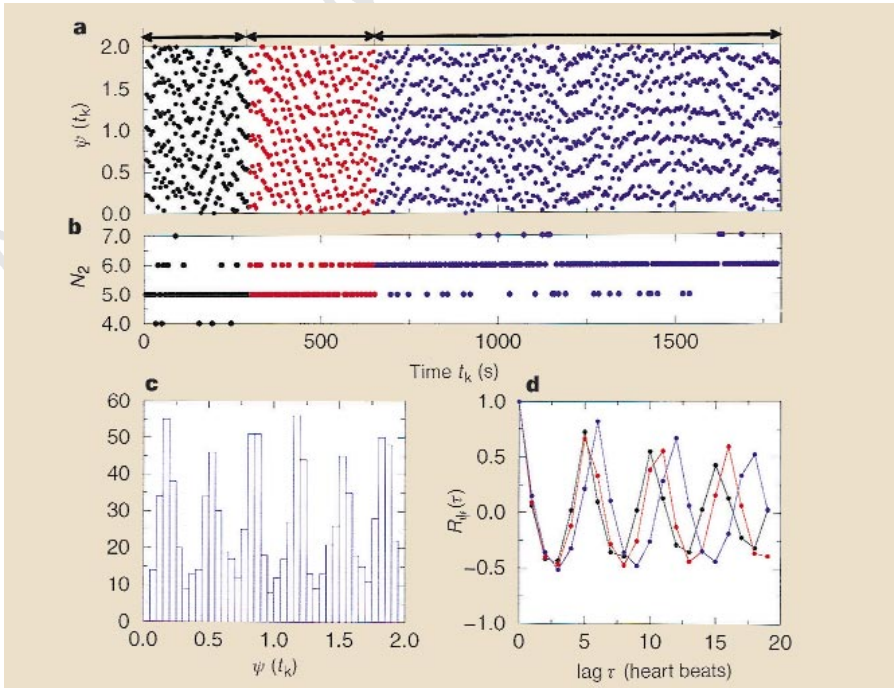


Figure 1 Analysis of cardiorespiratory cycles. **a**, Cardiorespiratory synchrogram, showing the transition (red) from 5:2 frequency locking (black) to 3:1 phase locking (blue). Each point shows the normalized relative phase of a heartbeat within two adjacent respiratory cycles $\psi(t_k) = (\phi_i(t_k) \bmod 4\pi) / 2\pi$. **b**, Number of heartbeats within two adjacent respiratory cycles. **c**, Histogram of phases. The six horizontal stripes in the blue region of the CRS result in six well-pronounced peaks in the distribution of phases. **d**, Autocorrelation function of phases $R_\psi(\tau) = \sum_i (\psi(t_k) - \langle \psi \rangle) (\psi(t_{k+\tau}) - \langle \psi \rangle) / \sum_i (\psi(t_k) - \langle \psi \rangle)^2$. The coloured curves correspond to respective regions.

1. Mackey, M. C. & Glass, L. *From Clocks to Chaos: The Rhythms of Life* Ch. 7 (Univ. Press, Princeton, 1988).
2. Stratonovich, R. L. *Topics in the Theory of Random Noise* (Gordon and Breach, New York, 1963).
3. Blekhnman, I. I. *Synchronization in Science and Technology* (ASME, New York, 1988).
4. Strogatz, S. H. & Steward, I. *Sci. Am.* **269**, 102–109 (1993).
5. Ludwig, C. *Arch. Anat. Physiol.* **13**, 242–302 (1847).
6. Anrep, G. V., Pascual, W., Rössler, R. *Proc. R. Soc. Lond. B* **119**, 191–230 (1936).
7. Eckberg, D. L., Kifle, Y. T., Roberts, V. L. *J. Physiol., Lond.* **304**, 489–502 (1980).
8. Pessenhofer, H. & Kenner, T. *Pflügers Arch.* **355**, 77–83 (1975).
9. Rosenblum, M. G., Pikovsky, A. S., Kurths, J. *Phys. Rev. Lett.* **76**, 1804–1807 (1996).
10. Pikovsky, A. S., Rosenblum, M. G., Osipov, G. V., Kurths, J. *Physica D* **104**, 219–238 (1997).

Teasing with winsome heresy

The Ascent of Science

by Brian Silver

Oxford University Press: 1998. Pp. 534.

\$35, £25

John Maddox

The Ascent of Science is written for *l'homme moyen sensuel*, or "HMS" as Brian Silver calls him. But who is this *homme*? Not quite "the man on the Clapham omnibus", as the southern English are fond of saying. Silver says his chosen reader has forgotten what little science he learned at school, is "more streetwise than the average scientist", and that, half the time, "he" means "she". In reality, he has written an account of the emergence of ideas in science for people who read broadsheet newspapers, not tabloids. So one should probably translate "*sensuel*" as "thinking" rather than "sensual".

Silver was professor of physical chemistry at the Technion Israel Institute of Technology until his death a few months ago. It is to be hoped that that sad circumstance does not mean his book fails to win the attention it deserves. For reviewers, the circumstance is an embarrassment: hortatory advice about doing better next time is sadly beside the point.

There are only a few complaints to make. Silver is right to remind us of the importance of the Enlightenment (call it the eighteenth century) in making science respectable, and of the backlash, led by Goethe at the beginning of the following century, making science romantic. But one can (and perhaps should) cavil at the opinion that Michael Faraday and James Clerk Maxwell shrank from the idea of the electron because of Goethe's influence, transmitted to the former by Humphry Davy (egged on by Coleridge) and then absorbed by Faraday's fan Maxwell. One looks (in vain) for chapter and verse.

Unsurprisingly, Silver, a physical chemist, is good on the molecular basis of thermodynamics. His HMS will get a feeling for the way molecular motion engenders pressure in a gas, and for the distinction between the collisions of a molecule whose only attribute is mass and the collisions of a person fighting his or her way through the crowd at Grand Central Station, who usually has a goal. That entropy should be linked with probability rings out as clearly as a bell.

Yet Silver also teases his readers unhelpfully, with questions such as "Is determinism dead?". It is easy enough to set up Pierre-Simon Laplace as the exponent of the ultra-Newtonian view that the future behaviour of the Universe and its contents could be calculated if only there were big enough machines, and to follow that with a

(superbly clear) account of chaos. The trouble is that the reality of chaotic motion is often taken as a proof (to HMS) that causality does not apply. Nothing could be further from the truth, of course, and HMS should be told so explicitly.

Much the same (but more) is true of Silver's treatment of quantum mechanics, which is winsome heresy. He lists seven "paradoxes", from Louis de Broglie's wave-particle duality through Schrödinger's cat to the Einstein, Podolski and Rosen experiment and its ramifications. And the conclusion? There "may be things beyond our ken and even our capacity to ken". Many HMSs will be content with that, but mistakenly.

The problem of quantum mechanics is not like that at all. The "commonsense" that Silver asks should be applied to quantum phenomena is a creature of our human senses, as refined by natural selection so far. It is no more surprising that we cannot sense the wave properties of electrons than that the reality of atoms has been estab-

lished only by a long and patient chain of inference. (Even the recent evidence of the scanning tunnelling microscope can be laughed off as artefactual, albeit with difficulty.) So why not accept that electrons (and other particles) are entities that have both wave and particle properties, according to the circumstances? De Broglie did not erect a paradox, but described the real world on a small scale.

Young's experiment (two slits and an interference pattern) is equally inaccessible to commonsense, but Silver goes too far in writing that "it cuts away at our understanding of the way the Universe is". Rather, it is the way the Universe is. Then to cite Richard Feynman as believing that this experiment is "the problem in quantum mechanics" (his italics) is unfair; Feynman meant merely that the double-slit problem contains all the commonsense difficulties of the Einstein, Podolski and Rosen experiment. Feynman's own way of doing quantum mechanics neatly resolves the difficulty by allowing a wave-



Mesmerized by meteors

Roberta J. M. Olson and Jay M. Pasachoff look back to what they call the golden age of British astronomy and art in *Fire in the Sky: Comets and Meteors, the Decisive Centuries, in British Art and Science* (Cambridge University Press, £50, \$74.95). They say that a greater number and variety of paintings and drawings of comets and meteors were produced in Britain during the eighteenth and nineteenth centuries than in other Western countries. The Smyths, father and son, typified the widespread interest

in astronomy at the time. William Henry Smyth was a keen amateur astronomer and an officer in the Royal Navy. His son, Charles Piazzi Smyth, became Astronomer Royal for Scotland in 1845 at the young age of 26, and was a prolific painter of astronomical phenomena. The picture above is his watercolour of Coggia's comet over Edinburgh in 1874. The book's authors say that Charles Piazzi was one of the first to champion the idea of placing observatories on mountains.

particle to follow all possible trajectories.

Silver's overall conclusion is in line with what he says of quantum mechanics: "...the attempt to understand the basic nature of reality may well be a losing game." He cites Gödel's theorem by way of support. All too many of the *HMSs* will be delighted with this nihilist message. □

Sir John Maddox is at 9 Pitt Street, London W8 4NX, UK.

Engineering branch

Design in Nature: Learning from Trees

by Claus Mattheck

Springer: 1998. Pp. 276. \$44.95, £30 (pbk)

Julian Vincent

Claus Mattheck grew up in East Germany, where the regime stole his history and his trust. He was a political prisoner for several years after trying to escape (all his East German equipment broke down and failed him; only a West German watch was reliable). Truth and trust are very important to him, and he seeks them in his teachers and friends, the trees.

I spent a happy day with him wandering around the Black Forest near Karlsruhe, hitting trees with a plastic hammer and diagnosing their structural design. I learned to see the lines of force stretching upwards from the roots and to detect the presence of a dead branch by listening to the echo in the trunk.

Later, in his laboratory in Karlsruhe, I saw how he applied his training as a physicist to produce a simple and compelling view of natural structures. By adapting techniques of finite element analysis, a standard tool used by structural engineers to analyse the balance of forces in complex shapes, Mattheck 'grows' shapes in the computer, making them spread the loads as evenly as possible. From simple beams and struts he derives the shapes of nature by erosion and accretion. This would be a pleasant enough pastime, but the method has great utility.

His *habilitation* was in fracture mechanics, so he was well placed to see that the shapes of nature, by eliminating self weight and stress concentrations, represent optimized solutions for engineering design. Throw away the design rulebook. Adopting the rules of nature using the 'soft kill option' (another of Mattheck's little jokes; he is a keen bear hunter, and follows them all day on foot with dogs) makes possible the design of screw threads, shells, levers and shafts that can support greater loads and are less prone to fatigue.

Mattheck has been developing these ideas over the past ten years in a number of publications alongside idiosyncratic illustrations, which his wife taught him to draw, and a multiplicity of photographs and computer

Marrow of the skull bears fruit

The corpus callosum of the human brain, from the atlas of Louis

Achille Foville's *Traite complet* (1844). For

centuries, so little was known about the true structure of this band

of white fibres connecting the cerebral hemispheres that Steno, in 1669,

confessed that "a man of a tolerable Genius may say about it whatever he pleases";

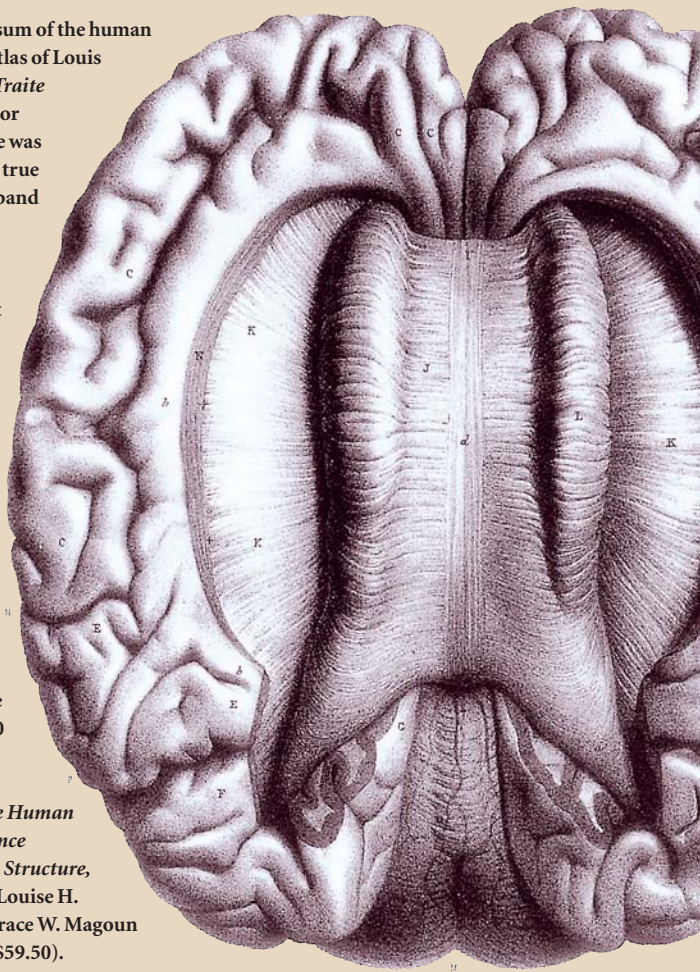
indeed, five years earlier, Willis had proposed that it housed the imagination. The plate is one of 300

archival illustrations in

Discoveries in the Human Brain: Neuroscience

Prehistory, Brain Structure, and Function by Louise H.

Marshall and Horace W. Magoun (Humana Press, \$59.50).



models. Much of this is in this book.

He summarizes the computer models he uses, and the reasons for using them, and applies them to growing, damaged and diseased trees, then to bone, claws, thorns, shell structures and bracing. Finally, he applies his methods to the design of a variety of engineering structures. I recommend this book to biologists and engineers alike. □

Julian Vincent is at the Centre for Biomimetics, University of Reading, Reading RG6 2AT, Berkshire, UK.

Decisive action

Sources of Power: How People Make Decisions

by Gary Klein

MIT Press: 1998. Pp. 330. \$40, £33.95

Valerie M. Chase

Like Edwin Hutchins' recent *Cognition in the Wild*, Gary Klein's *Sources of Power* examines how experts make decisions in real-world environments where time is short and stakes are high. Klein, the owner of a consulting company that advises organizations such as the military, hospitals and firefighting units, sets

out to discover "how people do so well under [such] difficult conditions".

The book's title refers to the tools that can be used to make decisions. To the list of traditional methods such as cost-benefit analysis, Klein adds tools that look less 'rational' but in fact serve experts well. These include 'the power of intuition', 'mental simulation' and 'metaphor'. Although Klein does not express his views precisely in the following terms, he seems to espouse Herbert Simon's brand of bounded rationality, the central idea of which is 'satisficing'. A satisficer sets an aspiration level — for instance, to hire one of the top 10% of job candidates rather than the single best one — and then applies heuristic strategies for meeting it. Satisficing allows people to make decisions that would require unmanageable amounts of information search and computation to make optimally.

The experts Klein has studied often make life-or-death decisions in a matter of minutes on the basis of incomplete, shifting information. Yet time pressure does not seem to compromise the quality of their decisions. In one study, Klein found that the quality of chess masters' moves was the same under blitz conditions (6 seconds per move) as under normal conditions (145 seconds

per move). Why? The first move they considered was disproportionately likely to be a good one. Klein argues that the more experience a decision-maker has in a domain, the better he or she is at satisficing in that domain. So the training programmes designed by Klein's company tend to involve teaching novices how to satisfice like experts rather than using rational decision-making methods.

Klein tries to account for expert decision-making with the 'recognition-primed decision' (RPD) model. Its first stage amounts to categorization of the situation. As well as a default course of action, this categorization activates 'expectancies' that help the decision-maker to detect anomalies. If there is a glitch, he or she can turn to mental simulation or analogical reasoning to tailor a new course of action.

Unfortunately, because Klein does not explain in detail how he and his colleagues analysed the interviews that serve as their primary data source, some of the quantitative findings reported in the book are difficult to interpret. Moreover, his efforts to ground the RPD model in basic research in psychology are spotty and idiosyncratic, which might irk some knowledgeable readers. The RPD model does not elucidate anything about naturalistic decision-making that cannot be gleaned from the book's numerous (and occasionally fascinating) case studies.

Why is satisficing not generally advocated by decision theorists? Part of the answer is that decisions based on this method are difficult to justify after the fact. Although many organizations train their people to use transparent, rational procedures, experienced decision-makers tend not to apply them. Decisions that look wrong in retrospect are often blamed on deviations from rational methods.

In July 1988, for example, during the Iran-Iraq war, the crew of the USS *Vincennes* shot down what turned out to be a commercial aircraft in the Persian Gulf in the belief that it was an Iranian fighter plane. One decision analyst later argued that expectancy bias distorted the commander and crew's interpretation of the information about the plane's speed and altitude. However, as Klein observes, if the plane had turned out to be a fighter, the same analyst could have blamed the error on base-rate neglect: in the previous month, most of the aircraft identified by the US forces in the Middle East had been Iranian military planes.

In hindsight, the decision to shoot was clearly an error, but, as reconstructed by Klein, it seems to have been reasonable given the information available.

Klein has amassed an impressive quantity and range of evidence that erodes the myth of the expert decision-maker who behaves according to classical rational models, and

he suggests that traditional definitions of both rationality and expertise need to be re-examined. □

Valerie M. Chase is at the Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Lantzeallee 94, 14195 Berlin, Germany.

Benevolent nuclei

Tracks to Innovation

by Robert L. Fleischer

Springer: 1998. Pp. 193. \$49.95, £37.50

Robert W. Cahn

The best book titles leave the potential reader intrigued and prone to curiosity. The title of Robert L. Fleischer's book is loaded with meaning. The overt subject matter is the tracks left in insulating materials by high-energy subatomic projectiles. Improbable though it might seem, the study of such tracks has led to a range of profitable industrial innovations (as well as to scientific insights of inestimable value).

Fleischer spent many years as a research physicist at the Corporate Research Center of General Electric in upper New York State. There his colleagues Buford Price and Robert Walker discovered, in 1961, that fission fragments from uranium leave behind damage traces, known as 'tracks', in slivers of mica, and that these tracks can be revealed by a chemical etchant, which penetrates along the tracks much faster than it eats away the bulk of the mica. As Fleischer points out, this serendipitous observation owed something to the fact that a colleague had squirrelled away a supply of high-quality synthetic mica made as a reserve in time of war, and another pair of investigators there had recently found out how to reveal dislocation lines in lithium fluoride by etching.

The following year, Price and Walker teamed up with Fleischer to pursue this discovery. Someone in the laboratory needed a controlled, ultraslow vacuum leak, and it occurred to the trio that etching the tracks right through a piece of mica could provide this, and indeed it did. The diameter of the hole increased with longer etching time, and the relationship was highly reproducible.

From that point on (and later with the help of an experienced nuclear physicist, Antonio Mogro-Campero), the trio combined fundamental studies with searches for applications. They examined such issues as the relative 'etchability' of tracks in different materials, crystalline or amorphous, the threshold dose to render a track etchable at all, the relative effectiveness of alpha particles and fission fragments of different energies (beta and gamma rays are ineffective), and the nature of 'track holes' under different irradiation and etching conditions. After three years of this, a confidential conference

was organized at General Electric involving 44 of the company's scientists and engineers from around the United States to explore the discovery's possible uses.

At about the same time, the first — highly profitable — application emerged through a chance meeting with a New York cancer researcher, who needed an ultrafilter to separate cancer cells and blood cells. It turned out that the company's own amorphous polycarbonate film was ideal for making a population of small, generally non-overlapping, perfectly round holes, from nanometres to several micrometres across, for use as ultrafilters. This process was soon commercialized by General Electric as 'Nuclepore', with an eventual annual profit of more than \$10 million. Now that the patent has expired and several companies have joined the fray, the worldwide demand for such filters for biomedical research, aerosol studies and various forms of analytical and laboratory use is about half a million square metres per year, with a value of \$50 million.

The team at General Electric perfected a bewildering array of scientific and technological uses including dosimetry of radon in houses, estimation of the time since rocks were last heated in mountain-building, and determination of plutonium distribution in an organism by a form of track-based autoradiography. The most profitable use of track technology (though the profit did not accrue to General Electric) was the use of rock-heating methodology by petroleum exploration companies: recent heating of a reservoir denatures the oil, and where such heating is deduced from a study of surface rocks, petroleum companies can save tens of millions of dollars that would otherwise have been used to drill an ineffective well.

The last chapter analyses concisely but profoundly, in the light of the 30-year adventure of track research, the uses of in-house scientific research in industry (which has steadily declined at General Electric, as it has in many companies worldwide). Fleischer points out that two observations of track etching were made at Harwell Laboratory in England before the first observations at General Electric, but led nowhere because, at the time, Harwell was not commercially orientated; General Electric was, and reaped the commercial benefit.

The important point is this: "A wealth of new opportunities can arise when people with different knowledge and interests come together." The diversity at General Electric ensured that the "boundaries of separate areas of knowledge" came together to take advantage of the work on track etching. The book provides a concise, informative, witty and easy-to-read account of this episode that is replete with lessons. □

Robert W. Cahn is in the Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK.

In retrospect chosen by Carl Djerassi

The Struggles of Albert Woods

by William Cooper
(1952)

On 24 March 1994, before a full house at the Royal Society of Literature, London, a panel under the chairmanship of the biologist Lewis Wolpert debated the issue: "Are writers too ignorant about science?". As one of the participants, I tried to argue that 'science-in-fiction' is a literary genre distinct from science fiction; that, although rare, it had an honourable tradition, especially in the United Kingdom; and that its most appropriate authorial practitioners should be insiders (meaning scientists), as the genre's most distinguishing features are accuracy and plausibility.

Admittedly, I had an ulterior motive, having already published two science-in-fiction novels. Still, Wolpert needed no convincing, and William Waldegrave, the British cabinet minister then responsible for science, also seemed persuaded (both had sampled my work), while the novelist Maggie Gee remained diplomatically neutral. It was the crime novelist P. D. James who disagreed: she was no scientist and knew little about science, she announced proudly, but had always known where to get scientific information if she needed it for her novels.

My hackles were raised subliminally. Because science-in-fiction deals not only with science, but more importantly with scientists too, I felt that a clansman can best describe a scientist's tribal culture and idiosyncratic behaviour. "Take C. P. Snow", I started to say, but then stopped. Everyone starts with Snow, notably with *The Search* (1934, revised 1958). "In fact, take William Cooper", I continued. "Take *The Struggles of Albert Woods*, which is far superior to *The Search*."

To my surprise, several people in the audience grinned, including a white-haired elderly gentleman with an officer's clipped moustache. Interpreting the grins as sniggers, I decided to defend both Cooper and my impeccable judgement by pointing out that, whereas Snow had no sense of humour, Cooper's was abundant. Seeing more grins, I proceeded to outline some high points of Cooper's novel: how Albert Woods, who, after gaining a first-class honours degree in chemistry from a redbrick university, started research on the 'untypical Wurmer-Klaus Reaction' (to organic chemists, a clear allusion to the 'abnormal Reimer-Tiemann Reaction'); and how this was the speciality of Oxford's reader in experimental chemistry, 45-year-old F. R. Dibdin, a man obsessed with obtaining an F.R.S. after his name. (As Cooper pointed out, given the then-existing pool of British research scientists, the odds were a measly 20 to 1 of being elected a fellow of the 500-member Royal Society. "At the age of 30 the thought of being elected is inspiring; at the age of

45 the thought of not having been elected is agonising.")

Dibdin, the brightest star in the Wurmer-Klaus firmament, eliminates the potential competitor Woods by inviting him to join his Oxford group. Misnaming him 'Bowls', Dibdin promptly cuts him down to size. "You've made a very good start, but that doesn't mean you're bound to make a good finish." Dibdin, in Cooper's words, "had no capacity at all for doing experiments... but was most ingenious in inventing experiments for other people to do. Dibdin's school had more irons in the fire per man than any other school in the country." But Woods, a superb experimentalist, finishes his project within a year. In a scene hilarious as well as realistic, he presents Dibdin with his first journal manuscript. Instead of the expected perfunctory comments, Woods is confronted with a *fait accompli*:

"I see you've put your own name at the top of the paper, Mr. Woods." His eyes looked sad and thoughtful. "I always make it a matter of principle to put my name as well on every paper that comes out of the department."

"Yours?" Albert said incredulously.

"Yes," said Dibdin, still sad and thoughtful. "I make it a matter of principle, Mr. Woods. And I like my name to come first — it makes it easier for purposes of identification." He rounded it off. "First come, first served."

And Cooper is not reluctant to bring in sex. To wit: "Nobody could ignore the imagination, enthusiasm, fervor, and most of all, palaver, with which Albert conducted his affairs with Thelma. The two-seater Morris-Cowley was a symbol of them all." There follows a playful description of the limitations of a Morris-Cowley front seat:

"Oh!" cried Thelma, with her eyes closed in excitement. The door of the motor car flew open and Albert fell on to the floor. Hotly he picked himself off the hand brake and gear lever and got back again. He was rough and powerful in his haste. "Oh," cried Thelma, this time with her eyes open. Albert felt her quivering beneath him and thought it was passion. His face was close to hers, his voice thick and breathless as he said: "Sometimes I'm terribly animal." Thelma did not speak. She was shaking with laughter.

Realizing that people might think that the novel was only a light romp through science, I pointed to the wit and timeliness of many of Cooper's observations: how to get elected to the Royal Society; how not to achieve a desired knighthood; how to attract graduate students and then to exploit them; how such suffering exploitation is perpetuated once the earlier victim reaches his own pinnacle of power; the

pros and cons of academia chasing after industrial money; "the bane of the organic synthesizer's life — being able to make something but not enough of it for the making to be a practical proposition"; and the ultimate expression of grantsmanship: "it was not the first time a scientist had confidently said he had done some experiments when he had not."

It is impressive how accurately Cooper describes the chemistry scene, considering that he graduated with third-class honours in physics from Christ's College, Cambridge, and then spent much of his life in the Civil Service Commission (where his path crossed Snow's, as it had at Cambridge).

Stylistically, the text is interspersed with charming Trollopian asides to "Dear Reader". For instance, in preface to a sexual scene, Cooper writes: "May I point out that anyone who knows he or she is bound to be shocked can save us all embarrassment by the simplest manoeuvre — just skip. Albert's life becomes perfectly respectable a little later, so if you feel the preliminary twinges of moral indignation, now is the time to take the remedy. It is better to skip than to burst."

Of course, I realized that if I continued to push the virtues of Cooper's novel, I would never get to my own, which would defeat the ultimate objective of my appearance on the panel. So I proceeded to graze on my own literary pasture. But later on, after re-reading the novel, I was struck by how much I still remembered after four decades. Only the charming, non-saccharine, Trollopian ending had faded from my memory.

After the debate, we mixed with the audience over drinks. That's when the moustachioed gentleman, whose earlier grin had derailed my intended focus on my first novel, *Cantor's Dilemma*, was introduced to me by another grinner. "Professor Djerassi, meet Harry Hoff." I had assumed that Harry Summerfield Hoff, alias William Cooper, had long been dead. Instead I had the rare pleasure of shaking the hand that had written *Albert Woods* some 40 years earlier. Carl Djerassi is in the Department of Chemistry, Stanford University, Stanford, California 94305-5080, USA. Web address: <http://www.djerassi.com>



Dendritic cells and the control of immunity

Jacques Banchereau & Ralph M. Steinman

B and T lymphocytes are the mediators of immunity, but their function is under the control of dendritic cells. Dendritic cells in the periphery capture and process antigens, express lymphocyte co-stimulatory molecules, migrate to lymphoid organs and secrete cytokines to initiate immune responses. They not only activate lymphocytes, they also tolerize T cells to antigens that are innate to the body (self-antigens), thereby minimizing autoimmune reactions. Once a neglected cell type, dendritic cells can now be readily obtained in sufficient quantities to allow molecular and cell biological analysis. With knowledge comes the realization that these cells are a powerful tool for manipulating the immune system.

Immunology has long been focused on antigens and lymphocytes, but the mere presence of these two parties does not always lead to immunity. A third party, the dendritic cell (DC) system of antigen-presenting cells (APCs), is the initiator and modulator of the immune response. First visualized as Langerhans cells (LCs) in the skin in 1868, the characterization of DCs began only 25 years ago. It was known that 'accessory' cells were necessary to generate a primary antibody response in culture, but it was only once DCs were identified and purified from contaminating lymphocytes and macrophages that their distinct function as APCs became apparent (Box 1).

DCs are efficient stimulators of B and T lymphocytes¹. B cells, the precursors of antibody-secreting cells, can directly recognize native antigen through their B-cell receptors. T lymphocytes, however, need the antigen to be processed and presented to them by an APC. The T-cell antigen receptors (TCRs) recognize fragments of antigens bound to molecules of the major histocompatibility complex (MHC) on the surface of an APC. The peptide-binding proteins are of two types, MHC class I and MHC class II, which stimulate cytotoxic T cells (CTLs) and helper T cells respectively. Intracellular antigens, cut into peptides in the cytosol of the APC, bind to MHC class I molecules and are recognized by CTLs, which, once activated, can directly kill a target cell. Extracellular antigens that have entered the endocytic pathway of the APC are processed there and generally presented by MHC class II molecules to T-helper cells, which, when turned on, have profound immune-regulatory effects.

Some *raison d'être* for a specialized DC system are now clear (Fig. 1). The initiation of T-cell immunity is rather demanding. Initially, peptides from infected cells located anywhere in the body must be found and recognized by T cells that circulate in the blood stream. The amounts of specific antigen-MHC complexes on tumours and infected cells are typically small (one hundred or less per cell), and must be recognized by rare T-cell clones (usually at

a frequency of 1/100,000 or less) through a TCR that has a low affinity (1 μ M or less). Moreover, infected cells and tumours frequently lack the co-stimulatory molecules that drive clonal expansion of the T cell, the production of cytokines, and development into killer cells. DCs provide a means of solving these challenges. Located in most tissues, DCs capture and process antigens, and display large amounts of MHC-peptide complexes at their surface (Fig. 1). They upregulate their co-stimulatory molecules and migrate to lymphoid organs, the spleen and the

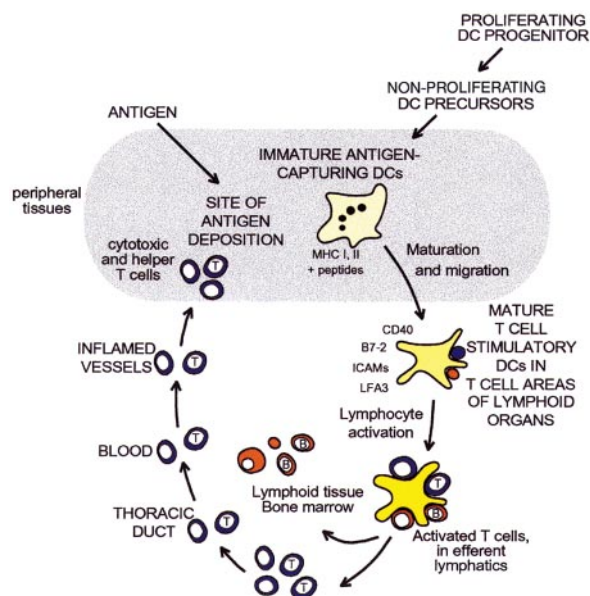


Figure 1 Afferent and efferent limbs of immunity that resolve several demands of antigen presentation *in vivo* (see text). Antigens are captured by DCs in peripheral tissues and processed to form MHC-peptide complexes. These immature DCs derive successively from proliferating progenitors and non-proliferating precursors, the latter not being fully committed to form DCs. As a consequence of antigen deposition and inflammation, DCs begin to mature, expressing molecules that will lead to binding and stimulation of T cells in the T-cell areas of lymphoid tissues. If the antigen has also been bound by B cells, then both B and T cells can cluster with DCs, as shown. After activation, T (blue) and B (orange) blasts leave the T-cell area. B blasts move to the lining of the intestine, the bone marrow, and other parts of the lymphoid tissue, such as the medulla of lymph node, with some becoming antibody-secreting plasma cells. T blasts leave the blood at the original site of antigen deposition, recognizing changes in the inflamed blood vessels and responding vigorously to cells that are presenting antigen. This limits the T-cell response to the site of microbial infection.

Box 1 Dendritic cells and the control of immunity

- Sentinels *in vivo*: *in situ* distribution to optimize antigen capture; migration into lymphoid organs to optimize clonal selection of rare, CD4⁺ and CD8⁺ T cells
- Initiators of immune responses: stimulation of quiescent, naive and memory, B and T lymphocytes
- Potency in stimulating T cells: capacity of small numbers of DCs and low levels of antigen to induce strong T-cell responses
- Inducers of tolerance: deletion of self-reactive thymocytes, and anergy of mature T cells

lymph nodes, where they liaise with and activate antigen-specific T cells (Fig. 1, right). All of these DC activities can be induced by infectious agents and inflammatory products, so that DCs are mobile sentinels that bring antigens to T cells and express co-stimulators for the induction of immunity.

Not only should the immune system attack that which is foreign or aberrant, it should also leave alone that which is neither to avoid autoimmunity. And again, DCs play a vital part. Normally, mature T cells do not respond to self-peptides that are presented to them: during their development only those T cells that have no, or low, affinity towards the peptide antigens present in the thymus are allowed to mature and to enter the circulation. But not all self-antigens are represented in the thymus. What prevents T cells from being activated by for example brain- or pancreas-derived self-peptides, an event that could ultimately result in the development of multiple sclerosis or diabetes? We shall discuss some evidence that DCs can tolerize the T-cell repertoire to self-antigens, both in the thymus and in the periphery.

Until recently, the paucity of DCs and the lack of specific markers had impeded research. Both obstacles are being overcome, and sufficient DCs can now readily be prepared from progenitors²⁻¹⁰: bone-marrow cultures stimulated with cytokines, and mouse cell lines cultured in GM-CSF (granulocyte-macrophage colony stimulating factor) and interleukin (IL)-4, are both tricks of the trade used to obtain reasonable numbers of DCs. DCs derived from human blood monocytes that have been nurtured in GM-CSF and IL-4, followed by maturation in a monocyte-conditioned medium, are the most potent APCs known^{8,9}. These DCs have many features

of primary DCs, including the expression of molecules that enhance antigen capture and selective receptors that guide DCs to and from several sites in the body (Fig. 2). New DC products are being rapidly identified by screening of complementary DNAs¹¹⁻¹³ that will provide more markers for identifying, characterizing and targeting DCs. We shall first review the features of mature DCs and their capacity to stimulate T cells, then discuss the immature antigen-capturing DCs, focusing on their migration to the lymphoid tissues and their development from progenitors; we shall consider their role in B-cell responses and the induction of tolerance and will finish by discussing their role in clinical immunology and pointing out the next challenges.

Features of mature dendritic cells

No other blood cell exhibits the shape and motility that give rise to the term 'dendritic' cell. *In situ*, as in the skin, airways and lymphoid organs, DCs are stellate (Fig. 3a). When isolated and spun onto slides, DCs display many fine dendrites (Fig. 3b). When looked at with an electron microscope, the processes are long (>10 µm) and thin, either spiny or sheet-like (Fig. 3c). When alive and viewed by phase-contrast microscopy, DCs extend large, delicate processes or veils in many directions from the cell body (Fig. 3d). These bend, retract, and re-extend in a non-polarized fashion for a day or more. Actin cables are scarce¹⁴. The shape and motility of DCs fit their functions, which are to capture antigens and select antigen-specific T cells.

Terminally differentiated or mature DCs can readily prime T cells (Box 1). Once activated by DCs, these T cells can complete the

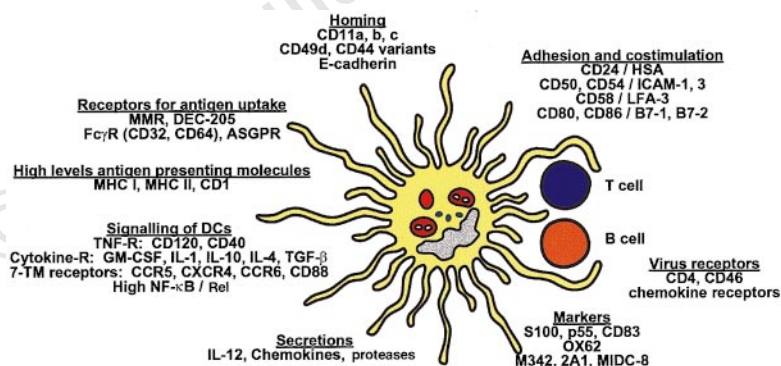


Figure 2 Some features of DCs, including DCs expanded *ex vivo* from precursors. Many monoclonals to antigens that are not rigorously DC-specific, nor well understood, are still useful in identifying DCs. These are CD83, an immunoglobulin-superfamily member; p55, a presumptive actin-binding protein; S100b,

a calcium-binding cytosolic protein; DEC-205, a multilectin in DCs and many epithelia; antibodies called M342, 2A1 and MIDC-8 which recognize antigens in intracellular granules of mouse DCs; and OX62 on rat DCs and some T cells.

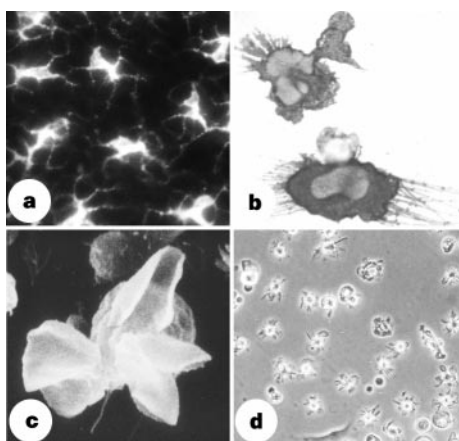


Figure 3 The unusual shapes of DCs. **a**, DCs in a sheet of epidermis (MHC II stain); **b**, in cytopins stained for surface MHC II; **c**, by scanning electron microscopy; **d**, in the live state by phase-contrast microscopy.

immune response by interacting with other cells, such as B cells for antibody formation, macrophages for cytokine release, and targets for lysis. Immature DCs, on the other hand, are less potent initiators of immunity but specialize in capturing and processing antigens to form MHC peptide complexes. Thus, two key functions of DCs segregate in time: they first handle antigens and then, as mature DCs a day or more later, stimulate T cells.

In vitro or *in vivo*, only few DCs are necessary to provoke a strong T-cell response. *In vitro*, DCs can induce a so-called mixed leukocyte reaction (MLR), a model for graft rejection. Leukocytes from one individual, the potential transplant donor, are mixed with T cells from the responder or graft recipient. If donor and recipient are mismatched at the MHC, the T cells begin to proliferate, release cytokines and become CTLs. Normally, the MLR is carried out with equal numbers of stimulators and responders, but only one DC is necessary to turn on 100–3,000 T cells. These early observations introduced the notion that there are cells, DCs, specialized to initiate immunity. Now it is clear that DCs prime T cells not only to mismatched MHC, but to a range of foreign proteins, from superantigens, the microbial proteins that bind directly to MHC molecules without prior processing¹⁵, to the larger world of more standard proteins that do require processing, including those from infectious agents^{16,17} and tumours^{18–21}. *In vivo*, immunity develops in lymphoid organs, where the DC–T-cell interaction can be seen for all major classes of T-cell ligands^{22–24}. DCs form clusters with antigen-specific T cells, creating a microenvironment in which immunity can develop^{22–24}.

So far, no one has been able to identify a single specific molecule to explain the efficacy of DCs in T-cell binding and activation, and the special effects of DCs seem solely to relate to quantitative aspects and their regulation. For example, MHC products and MHC–peptide complexes²⁵ are 10–100 times higher on DCs than on other APCs like B cells and monocytes. Mature DCs resist the suppressive effects of IL-10, but synthesize high levels of IL-12 (refs 26–28) that enhance both innate (natural killer cells) and acquired (B and T cells) immunity. DCs also express many accessory molecules that interact with receptors on T cells^{29,30} to enhance adhesion and signalling (co-stimulation), for example LFA-3/CD58, ICAM-1/CD54, B7-2/CD86. All these properties (MHC expression, secretion of IL-12 and the expression of co-stimulatory molecules) are upregulated within a day of exposure to many stresses and dangers, including microbial products.

Depending on the conditions, DCs can stimulate the outgrowth and activation of a variety of T cells, which affect the immune

response differently. They can persuade CTLs, which express the accessory molecule CD8 and hence interact with MHC class I bearing cells, to proliferate vigorously, which is unusual for CD8⁺ T cells^{31,32}. CD4-expressing T-helper cells, on the other hand, scrutinize cells that express MHC class II molecules. In the presence of mature DCs and of the IL-12 they produce^{26–28}, these T cells turn into interferon- γ (IFN- γ)-producing Th1 cells. IFN- γ activates the antimicrobial activities of macrophages and, together with IL-12, it promotes the differentiation of T cells into killer cells. So the capacity of DCs to produce IL-12 and Th1 cells will lead to microbial resistance. With IL-4, however, DCs induce T cells to differentiate into Th2 cells which secrete IL-5 and IL-4. These cytokines activate eosinophils and help B cells to make the appropriate antibodies, respectively.

The communication between DCs and T cells seems to be a dialogue rather than a monologue in which the DCs respond to T cells as well. CD40 (ref. 33) and the newly described TRANCE/RANK receptor^{34,35} on DCs are ligated by the TNF (tumour-necrosis factor) family of proteins expressed on activated and memory T cells: this leads to increased DC survival^{33,34} and, in the case of CD40, upregulation of CD80 and CD86 (ref. 33), secretion of IL-12 (refs 26, 27) and release of chemokines such as IL-8, MIP-1 α and β ³³.

Immature antigen-capturing dendritic cells

In most tissues, DCs are present in a so-called ‘immature’ state, unable to stimulate T cells. Although these DCs lack the requisite accessory signals for T-cell activation, such as CD40, CD54 and CD86, they are extremely well equipped to capture antigens and—a key event in the induction of immunity—antigens are able to induce full maturation and mobilization of DCs.

The sentinel position of immature DCs stand out when the skin surface, or epidermis, is labelled for DC molecules (Fig. 3a). Humans have about 10⁹ epidermal LCs, the immature dendritic cells of the skin that are located above the basal layer of proliferating keratinocytes. Freshly isolated LCs are weak T-cell stimulators, have few MHC- and accessory- molecules, but many antigen-capturing Fc γ and Fc ϵ receptors. This phenotype changes dramatically within a day of culture (Fig. 4): the cells undergo extensive transformation, antigen-capturing devices disappear, and T-cell stimulatory functions increase. When a skin patch is explanted and the LCs are challenged with an antigen, the migration of mature DCs into the culture medium can be observed³⁶. The situation is similar *in vivo*. When they encounter a powerful immunological stimulus, for example a contact allergen or a transplant, most of the LCs from

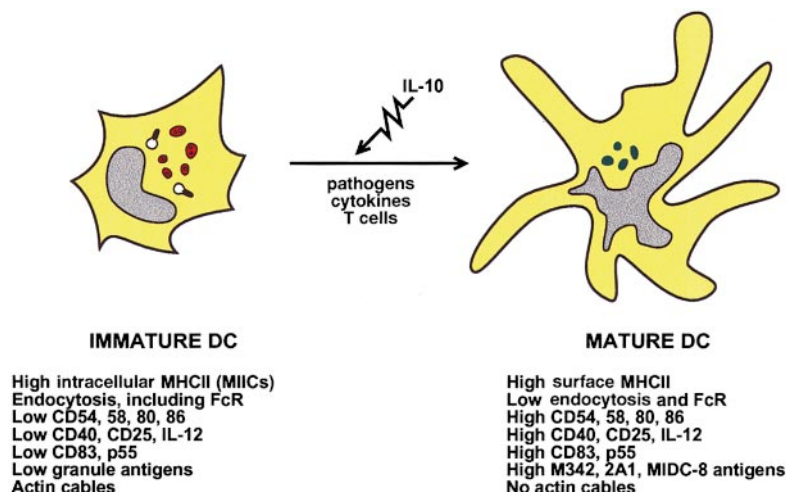


Figure 4 Features that change during DC maturation. Immature DCs take up antigen in several ways: that is, phagocytosis, macropinocytosis or adsorptive pinocytosis. An example of a pathogenic molecule that will induce maturation is

lipopolysaccharide (LPS); TNF α , GM-CSF are examples of cytokines, and CD40L is an example of a T-cell ligand that binds CD40 on DCs. IL-10 can inhibit maturation.

the epidermis mature and move into dermal lymphatics in search of antigen-specific T cells. Small numbers of antigen-capturing DCs can also be isolated from blood, lung, spleen, heart, kidney, and the B- and T-cell areas of tonsils; these cells lack LC-specific markers (E-cadherin, Birbeck granules, Lag-1), but they also acquire accessory molecules within 1–2 days of culture, before any encounter with T cells. In summary, a pool of cells in our bodies seems to be committed to mature into potent antigen-presenting cells.

What is it that makes a DC such a good APC? Immature DCs have several features that allow them to capture antigen. First, they can take up particles and microbes by phagocytosis^{16,17,37,38}. Second, they can form large pinocytotic vesicles in which extracellular fluid and solutes are sampled, a process called macropinocytosis⁷. And third, they express receptors that mediate adsorptive endocytosis, including C-type lectin receptors like the macrophage mannose receptor⁷ and DEC-205 (ref. 39), as well as Fcγ and Fcε receptors⁶. Macropinocytosis and receptor-mediated antigen uptake make antigen presentation so efficient that picomolar and nanomolar concentrations of antigen suffice⁷, much less than the micromolar levels typically employed by other APCs. However, once the DC has captured an antigen, which also can provide a signal to mature, its skills to capture antigens rapidly decline, and the time has come to assemble antigen–MHC class II complexes.

The antigen enters the endocytic pathway of the cell. In macrophages most of the protein substrate is directed to the lysosomes, an organelle with only few MHC class II molecules, where the antigen is fully digested into amino acids. Not in DCs; the DC is able to produce large amounts of MHC class II–peptide complexes at a single brief stage of its life. Much of this success may be due to specialized, MHC class II-rich compartments (MIICs) that are abundant in immature DCs^{6,14,40,41}. MIICs are late-endosomal structures that contain the HLA-DM or H-2M products, which enhance and edit peptide binding to MHC class II molecules. During maturation of DCs, MIICs convert to non-lysosomal vesicles that discharge their MHC–peptide complexes to the surface^{41,42} (Fig. 5). Immature DCs have been compared to idling motors in neutral gear, constantly degrading MHC class II molecules in their MIICs. As soon as an antigen instructs the DCs to move into gear, fragments of antigen are loaded onto class II molecules and these complexes are sent to the cell surface, where they remain stable for days^{41,42}.

To generate cytotoxic killer cells, which have the capacity to eliminate infected cells and attack transplants and tumour cells, DCs

have to present antigenic peptides complexed to MHC class I molecules to CD8-expressing T cells. This is relatively straightforward if the DC is infected itself, with, for instance, influenza virus. The virus uses the cell's machinery to synthesize viral proteins, which are, like cellular proteins, degraded into peptides by the proteasome. A dedicated peptide transporter then translocates these peptides from the cytosol to the endoplasmic reticulum, where they bind to class I molecules. The peptide-loaded MHC class I complexes travel to the cell surface where they are displayed for scrutiny by T cells. It is less clear, however, how DCs could process and present antigens that have no access to the cytosol in an MHC class-I-restricted manner (for instance transplant- and tumour-derived antigens, or antigens from viruses that cannot infect DCs). Nevertheless, they can probably present peptides from non-replicating microbes³² and dying infected cells⁴³ by means of MHC class I as efficiently as if they were infected themselves. By processing dying cells, DCs may be able to 'cross-prime' or 'cross-tolerize' T cells to another cell's antigens or self-proteins, which could have major clinical implications, as we will discuss.

It is clear that maturation of DCs is crucial for the initiation of immunity. It can be influenced by a variety of factors, notably microbial and inflammatory products. Whole bacteria¹⁴, the microbial cell-wall component LPS⁷, and cytokines like IL-1, GM-CSF and TNF-α, all stimulate DC maturation, whereas IL-10 blocks it⁴⁴. Ceramide, which is induced by maturation signals, can shut down antigen capture by the DC⁴⁵. Mature DCs express high levels of the NF-κB family of transcriptional control proteins (Rel A/p65, Rel B, Rel C, p50, p52)⁴⁶ which regulate the expression of many genes encoding immune and inflammatory proteins. Signalling through the TNF-receptor family, for example TNF-R (CD120a/b), CD40 and TRANCE/RANK, results in activation of NF-κB. Therefore, to induce the immune response through activation of DCs, a pathogen or antigen may have to engage the signal transduction pathways of the TNF-R family and TNF-R-associated factors (TRAFs).

Migration of dendritic cells *in vivo*

Upon activation, DCs travel to the lymphoid tissues such as the spleen and lymph nodes. There, DCs may complete their

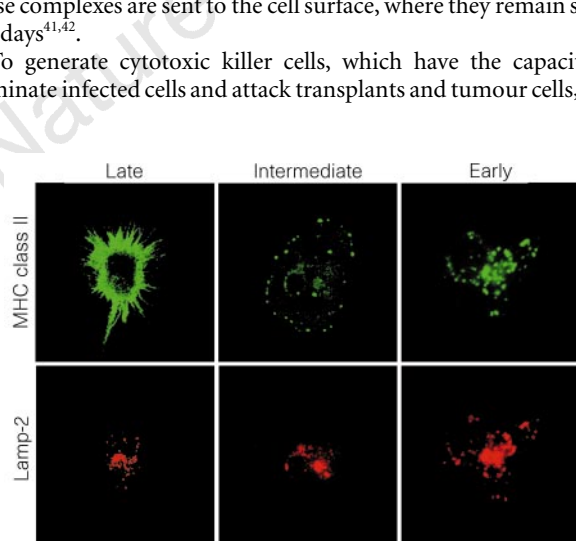


Figure 5 Intracellular MHC II-bearing compartments in immature, maturing and mature DCs. MHC II is shown in green and lysosomal membrane glycoprotein (Lamp-2) (or H-2M) in red. Early or immature DCs have numerous MHC II-rich lysosomes (right); maturing DCs contain MHC II-positive, non-lysosomal vacuoles (green, middle panels) that probably bring large amounts of MHC II to the surface of late or mature DCs (left). Micrographs courtesy of S. Turley and I. Mellman.

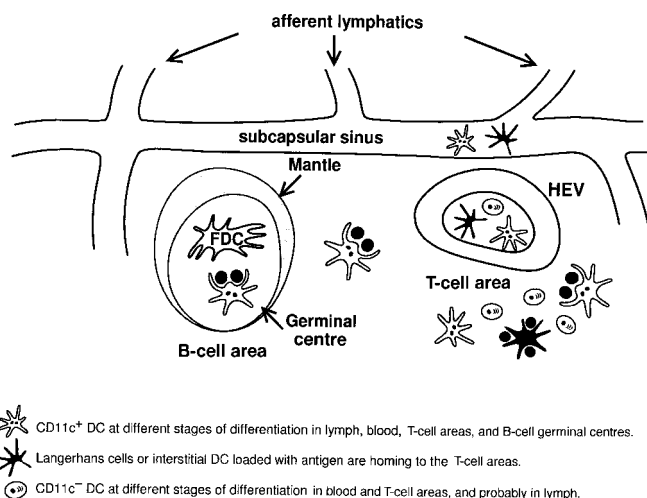


Figure 6 Distinct subsets of DCs in lymphoid organs, possibly derived from separate pathways of development (see text). DCs arrive through afferent lymphatics and HEVs and include cells that are on patrol, or have left peripheral tissues in response to a stimulus, or act as precursors to lymphoid DCs. All DCs may initially enter the T-cell regions to mature into what are called 'interdigitating' cells by electron microscopy. The activated B-cell follicles, with germinal centres, have two cell types: FDCs and CD11c⁺ DCs, which present respectively native antigens to B cells and processed antigens to T-helper cells. HEV, high endothelial venule.

maturation⁴⁷, attract T and B cells by releasing chemokines¹¹ and maintain the viability of recirculating T lymphocytes⁴⁸.

Even in the absence of invading pathogens, a fraction of the DC population seems to move around. These cells are always present in afferent lymph that access the T-cell areas (Fig. 6), but not in efferent lymph, indicating that most of the migrating DCs die after their arrival in lymphoid tissues. In blood, there are two subsets of precursor DCs, one expressing and one lacking CD11c (ref. 49). Both subsets can enter lymphoid organs at high endothelial venules by virtue of CD49d β 1 integrins⁵⁰ possibly with different final destinations such as the B-cell (CD11c⁺) and T-cell (CD11c⁻) areas (Fig. 6). DCs along liver sinusoids move in hepatic lymphatics to coeliac lymph nodes^{24,51}, and DCs from intestine migrate to mesenteric lymph nodes. The DCs that migrate in the steady state may replenish immature populations or may be on patrol to identify invaders, like other white blood cells.

Not every pathogen or antigen induces a strong T-cell response, but those that do can induce the mobilization and maturation of DCs. Inhaled viruses and bacteria mobilize DCs into airway epithelium, and this influx proceeds as rapidly as that of neutrophils⁵². An injection of LPS causes a massive egress of DCs from skin, heart, kidney and intestine^{53,54}, as well as the maturation and movement of spleen DCs from the marginal zone into the T-cell areas⁵⁵. Application of a fluorescent contact allergen to the skin induces DC migration, and a day later fluorescent DCs can be found in the draining nodes. Following transplantation of heart and skin, DCs crawl out of the graft and make their way to lymphoid tissues. When access is prevented by disconnecting the afferent lymphatics, no immune response to skin transplants and contact allergens develops.

How dendritic cells know where to go is largely unknown. LPS⁵³ stimulates a variety of cells to produce cytokines and chemokines, for example GM-CSF, TNF- α , IL-1, MIP-1 α and β . These products are known to modulate DC movement and maturation. Damage to cytokine-rich cells, such as keratinocytes or mast cells, induced by an antigen could result in the release of preformed mediators like GM-CSF, TNF- α and IL-4. The current focus, however, is on seven-transmembrane-spanning G-protein-coupled receptors, an ever-growing family of proteins that includes receptors for calcitonin-receptor-related peptide (which is found in nerve endings that abut on skin LCs⁵⁶), C5a and chemokines^{13,57}. These peptides and their receptors may mediate many of the steps required for the migration and targeting of DCs: egress from a tissue, directed movement or chemotaxis, and maintenance of viability.

Dendritic cell development

There seem to be several pathways to generate DCs. We mentioned earlier that blood monocytes give rise to DCs when cultured with the appropriate cytokines⁶⁻⁹. The DC progenitors are also present in bone marrow: a small CD34⁺ subset of haematopoietic progenitors gives rise to all blood cells and DCs. The DC progeny is likely to colonize most tissues *in vivo* as immature non-dividing cells. Several cytokines contribute to their growth and differentiation. Both c-Kit ligand and Flt-3 ligand are transmembrane proteins on stromal cells that bind to tyrosine-kinase receptors and sustain DC progenitors⁵⁸; administration of Flt-3 ligand *in vivo* stimulates the outgrowth of functional DCs¹⁰. GM-CSF and IL-3 (refs 2, 59), products of activated T cells and other cells, also enhance DC differentiation, whereas macrophage (M)-CSF favours differentiation of the precursors into macrophages. TNF and CD40L^{3,33,60} block the granulocyte-differentiation pathway and stimulate the final maturation of DCs.

Cells that express the marker CD34 contain progenitors for two discrete DC populations: the epidermal LCs and dermal or interstitial type DCs^{61,62}. One functional difference is that only interstitial-type DCs directly stimulate naive B cells to make antibodies⁶³. The LC progenitor expresses CLA (a ligand for E-selectin and a skin-homing molecule) and lacks CD14 (a marker that is abundant on

monocytes) and cannot form macrophages. In contrast, the dermal DC progenitors lack CLA, give rise to CD14-positive cells that resemble monocytes, and can form either macrophages in response to M-CSF or DCs in response to GM-CSF and TNF- α ^{61,64}. Mice deficient in TGF- β mice have no epidermal LCs but maintain DCs in other sites⁶⁵, which underscores the peculiarity of LCs.

A tentative but intriguing subset of DCs may be specialized in the induction of immune tolerance⁶⁶. It is proposed that these DCs arise from a progenitor that also yields T cells⁶⁷. IL-3, not GM-CSF, drives the development of these lymphoid DCs, which lack myeloid markers (markers shared by macrophages and granulocytes such as CD11b, CD13 and CD33)⁵⁹. In humans, a distinct lymphoid DC precursor may have been identified; it expresses CD4, lacks CD11c and looks like an antibody-forming plasma cell but develops into a DC in the presence of IL-3 and CD40L⁶⁸.

It is noteworthy that although mouse DC lines can be generated with the help of transforming viruses, there are no examples of spontaneous DC malignancy. In human Langerhans cell granulomatosis, originally called histiocytosis X, typical LCs and eosinophils accumulate. The cause of these granulomas is unknown, but they appear to be clonal in origin and create symptoms as a result of their position in critical sites rather than by overt malignant transformation⁶⁹.

Dendritic cells and B lymphocytes

DCs, famous for their T-cell-stimulatory properties, are now known to have major effects on B-cell growth and immunoglobulin secretion. B cells and DCs are both APCs and both are essential for antibody responses but for entirely different reasons (reviewed in ref. 1; Box 2); DCs activate and expand T-helper cells, which in turn induce B-cell growth and antibody production. But during this *ménage à trois*, there is more direct DC-B-cell dialogue as well (Fig. 1).

Naive B cells respond uniquely to the interstitial, non-LC type of DC^{63,70}, and by secretion of soluble factors⁷⁰, including IL-12, DCs stimulate the production of antibodies directly and the proliferation of B cells that have been stimulated by CD40L on activated T cells. DCs also orchestrate immunoglobulin class-switching of T-cell-activated B cells: IL-10 and TGF- β can induce secretion of IgA₁, but expression of IgA₂ appears to be strictly dependent on a direct interaction between the B cell and the DC⁷¹. This indicates that DCs are in control of mucosal immunity, and, in fact, DCs can be found in mucosal lymphoid tissues beneath antigen-transporting M cells⁷²⁻⁷⁴ and in T-cell areas.

Follicular dendritic cells, or FDCs, directly sustain the viability, growth and differentiation of activated B cells (Fig. 6). They also organize the primary B-cell follicles, as shown by the absence of FDCs and follicles in TNF- α -knockout mice. FDCs differ from ordinary DCs: they are not bone-marrow-derived⁷⁵, they lack the leukocyte marker CD45, and they display a unique set of molecules at their surface (reviewed in ref. 76), including all known complement receptors (CD11b, a long isoform of CD21 (ref. 77) and CD35). With their receptors for complement and Fc, FDCs capture antibody-antigen complexes and display whole complexes, rather than processed antigens, at their surface for long periods. FDCs are abundantly present within antigen-stimulated B-cell areas, or germinal centres. There, proliferating B cells (centroblasts) undergo

Box 2 Differences between dendritic cells and B cells as APCs

- DCs have higher levels of MHC II and accessory molecules
- DCs can make large amounts of IL-12
- DCs can internalize antigens by means of Fc and multilectin receptors; B cells have antigen-specific immunoglobulin receptors and inhibitory Fc γ R

somatic mutation, after which they stop dividing (centrocytes) and wait to be triggered by an immune complex on FDCs. B cells that recognize an immune complex with high affinity process the antigen and present it as peptide-MHC complexes to antigen-specific T cells. The T-B-cell interaction ensures the survival of these high-affinity B cells, while the non-stimulated low-affinity B cells apoptose and are phagocytosed by tingible body macrophages.

It is now known that germinal centres contain a second type of dendritic cell, the CD11c⁺ DC (Fig. 6). It also carries immune complexes, and is a much more powerful stimulator of T cells than the germinal centre B cells⁷⁸. This may be the DC that brings the antigens to the germinal centre and displays processed antigens to memory T cells.

Dendritic cells and T-cell tolerance

Most studies focus on the power of DCs to activate T cells, but before T cells encounter foreign antigens, the T-cell repertoire should be tolerized to self-antigens. This occurs in the thymus (central tolerance) by deletion of developing T cells, and in lymphoid organs (peripheral tolerance) probably by the induction of anergy or deletion of mature T cells. In both cases, the DC system that initiates immunity to foreign antigens also appears to tolerize T cells to self-antigens.

In the thymic medulla, DCs present self-antigens in the context of MHC molecules. Thymocytes that have too high an affinity for self-antigens are deleted (negative selection). If antigen-bearing DCs are directly injected into the developing thymus or fetal thymic organ cultures, reactive T cells are deleted. In the thymic cortex, macrophages digest large numbers of dying thymocytes that have failed to undergo positive selection. As these macrophages handle large amount of self-antigens, they seem ideally suited to delete auto-reactive T cells, yet they do not seem to do so: if MHC class II molecules are solely present on DCs in the medulla, negative selection ensues⁷⁹. If, on the other hand, MHC class II molecules are only expressed by cortical epithelium, and not by DCs in medulla, the propensity to autoimmunity increases⁸⁰, indicating that DCs in the medulla are responsible for the deletion of auto-reactive T cells. Also, *in vitro* it is the DC, and not the macrophage, that efficiently deletes auto-aggressive thymocytes⁸¹.

Recent studies point to an important role for DCs in the induction of peripheral tolerance as well. DCs can capture and present self-antigens that are exclusive to specialized tissues. For example, bone-marrow-derived APCs present peptides, which are derived from insulin-producing β -cells of the pancreas, to T cells in the draining lymph nodes^{82–84}; tolerance ensues, probably as a result of T-cell anergy or deletion^{83,84}. It has recently been shown that DCs present peptides from apoptotic cells⁴³. Accordingly, DCs may be able to present many self-antigens, derived from the normal turnover of somatic cells, to T cells and thus induce tolerance to self-proteins that have no access to the thymus.

What determines whether a DC turns the immune system on or off? Lymphoid DCs are long-lived cells and express very high levels of MHC-self peptide complexes²⁵, and maybe T cells become anergic or die in response to abundant and persistent antigens. Maybe distinct DCs are responsible for the different tasks: more resident lymphoid DCs induce tolerance to self, whereas migratory myeloid DCs, including LCs, are activated by foreign antigens in the periphery and move to lymphoid organs to initiate an immune response. Another possibility is that tolerance-inducing DCs are qualitatively different, perhaps expressing a death molecule like Fas ligand⁶⁶.

Dendritic cells in clinical immunology

Given their central role in controlling immunity, DCs are logical targets for many clinical situations that involve T cells: transplantation, allergy, autoimmune disease, resistance to infection and to tumours, immunodeficiency, and vaccines. In autoimmune diseases

such as psoriasis and rheumatoid arthritis, increased numbers and activation of DCs have been noted. DCs are important APCs in the lung, possibly contributing to allergy and asthma. In transplantation and contact allergy, DCs have been implicated in the induction of both immunity and tolerance. Here we concentrate on three areas where direct data on the role of DCs is available: infection with immunodeficiency viruses, resistance to tumours, and new approaches to vaccination.

Recent results paint a paradoxical picture in which DCs, instead of inducing host resistance, provide a safe haven for several viruses. Cells of the DC system may be hosting latent cytomegalovirus⁸⁵, and Kaposi's virus (KSHV) may likewise be sheltered in patients with multiple myeloma⁸⁶. For HIV-1 and measles, the consequences of DC infection are more overt: especially upon interacting with memory T cells and activated T cells, they sustain the production of many HIV-1, SIV and measles^{87–93} particles. Measles turn DCs into multinucleated cells, or syncytia, and suppresses dendritic-cell and T-cell function^{91,93}. HIV-1 and SIV also vigorously replicate in DC-derived syncytia *in vitro*^{90,94}, but immunosuppression is not apparent at this time. Syncytia are not just a sign of viral toxicity, as is often assumed, but are also true virus factories^{90–94}. *In vivo*, infected syncytia have been noted on the surfaces of mucosa-associated lymphoid tissue^{72,73}. These so-called lymphoepithelia contain numerous memory B and T cells, as well as DCs that are chronically exposed to maturation stimuli from the environment. But, rather than battling with the infection, mature DCs assist in its spreading by transmitting HIV-1 and SIV to T cells^{87,88,90,94}. FDCs in B-cell areas (Fig. 6) appear not to be infected with immunodeficiency virus, but they may play a dual role. By displaying virions complexed with antibody, they nevertheless can elicit resistance, especially B-cell memory, but at the same time they act as long-lived extracellular reservoirs of potentially infectious virus.

Many tumour components do not elicit an antigen-specific T-cell response in patients, which may be due to the absence of functional DCs in tumours. DCs that infiltrate colon and basal-cell skin cancers can lack CD80 and CD86 (ref. 95) and therefore have reduced T-cell stimulatory activity. Likewise, tumours may secrete factors, such as IL-10, TGF- β and vascular endothelial growth factor, that reduce DC development and function. Be that as it may, the more DCs infiltrate the tumour, the better the prognosis.

The immune repertoire carries tumour-reactive T cells, especially CTLs, but there is little evidence that these T cells are being activated *in vivo*. However, when tumour antigens are applied to DCs *ex vivo* and these DCs are then reinfused, specific immunity ensues. In animals this strategy can lead to protection against tumours and even a reduction in the size of established tumours^{96–98}, and at present similar studies are carried out in patients. Many vehicles for the delivery of tumour antigens to DCs are being considered: viral vectors, naked and plasmid DNA, RNA, liposomes with nucleic acid or protein, and tumour lysates, apoptotic cells and peptides. It is interesting that DCs appear to have a direct lytic potential on certain tumour targets as well⁹⁹.

Vaccine design has yet to target the DC system. DCs can readily elicit helper and killer T cells, antibodies and IL-12, and can operate at mucosal surfaces where protection is needed early during many infections. In contrast, many existing vaccines and adjuvants are weak stimulators of CD8⁺ T cells and Th1-type T cells. As some DCs appear to tone down the immune response, vaccines that target these DCs could also be used to induce tolerance, for example to allergens.

The classical approach to vaccination exploits attenuated forms of pathogens to elicit an immune response, and such attenuation is now more feasible using genetic manipulation and new vectors like avipox viruses. *In vitro*, DCs are the only cells that efficiently present inactivated virus³², and therefore the efficacy of the new generation of attenuated vaccines could be improved by specific targeting to DCs. The immune response can also be boosted by immunization

with DNA vaccines; even though the DNA is primarily expressed in weak APCs, like dermal and muscle cells, DNA vaccines can activate both CD4- and CD8-bearing cells. DCs isolated from vaccinated animals both express the vaccine DNA^{100,101} and present the corresponding peptides to specific T cells¹⁰¹. The frequency of transfected DCs is low and greater efficiency should prove valuable.

Upcoming challenges

We have only just started to appreciate the extraordinary life of dendritic cells and many questions remain unanswered. At the molecular level, there is a sizeable effort to screen expressed sequence tags from DCs to characterize their products. Regulatory molecules of the killer-inhibitor family¹⁰², of chemokines and chemokine receptors^{11,13,103}, and of proteinases¹² have now been identified. This ongoing search should uncover new lymphocyte-binding and antigen-processing molecules, secretory products, and transcription factors that may help to explain the specialized features of DCs.

At the physiological level, many features of DCs need to be understood in more detail to allow successful manipulation of the immune system. What signals lead to maturation and migration of DCs, how do they know where to go, and how do they reach the lymphatics, T-cell areas, and sometimes B-cell areas? What are the roles of the different DC subsets, including their effects on B cells and tolerance? Have DCs a larger repertoire of secretory products than currently appreciated, and are some of these, like IL-12, immune modulators? And, what determines whether a DC turns the immune system on or off?

At the clinical level, the efficacy of DNA vaccines and vectors may crucially depend upon their capacity to transduce DCs. Vaccines in general should improve if they elicit mobilization and maturation of DCs. Emerging evidence points towards a role for DCs in tolerance, and it will therefore be imperative to pursue DCs more vigorously in allergy, transplantation and autoimmunity. On the other hand, many microbes such as HIV-1 and measles may take advantage of the special features of DCs. Widely used immunomodulators like steroids may block DC maturation⁴⁷, and perhaps new drugs can be designed to modulate their functions. The results are awaited of clinical trials in which DCs are being used to elicit immunity to viral infections and tumours.

Although studies of pathogenesis have long focused on the core of immunology by identifying antigens and lymphocytes, all the recent evidence places a new emphasis on the role of DCs in the control of immunity. □

Jacques Banchereau is at the Baylor Institute for Immunology Research, Baylor Research Institute, 3409 Worth Street, Dallas, Texas 75246, USA; Ralph M. Steinman is in the Laboratory of Cellular Physiology and Immunology, Rockefeller University, 1230 New York Avenue, New York, New York 10021-6399, USA.

1. Steinman, R. M. In *Fundamental Immunology* (ed. Paul, W. E.) 4th edn (Lippincott-Raven, Philadelphia, in the press).
2. Inaba, K. et al. Generation of large numbers of dendritic cells from mouse bone marrow cultures supplemented with granulocyte/macrophage colony-stimulating factor. *J. Exp. Med.* **176**, 1693–1702 (1992).
3. Caux, C., Dezutter-Dambuyant, C., Schmitt, D. & Banchereau, J. GM-CSF and TNF- α cooperate in the generation of dendritic Langerhans cells. *Nature* **360**, 258–261 (1992).
4. Szabolcs, P., Moore, M. A. S. & Young, J. W. Expansion of immunostimulatory dendritic cells among the myeloid progeny of human CD34⁺ bone marrow precursors cultured with *c-kit* ligand, granulocyte-macrophage colony-stimulating factor, and TNF- α . *J. Immunol.* **154**, 5851–5861 (1995).
5. Romani, N. et al. Proliferating dendritic cell progenitors in human blood. *J. Exp. Med.* **180**, 83–93 (1994).
6. Sallusto, F. & Lanzavecchia, A. Efficient presentation of soluble antigen by cultured human dendritic cells is maintained by granulocyte/macrophage colony-stimulating factor plus interleukin 4 and downregulated by tumor necrosis factor α . *J. Exp. Med.* **179**, 1109–1118 (1994).
7. Sallusto, F. & Lanzavecchia, A. Dendritic cells use macropinocytosis and the mannose receptor to concentrate antigen in the MHC class II compartment. Downregulation by cytokines and bacterial products. *J. Exp. Med.* **182**, 389–400 (1995).
8. Romani, N. et al. Generation of mature dendritic cells from human blood: An improved method with special regard to clinical applicability. *J. Immunol. Meth.* **196**, 137–151 (1996).
9. Reddy, A., Sapp, M., Feldman, M., Subklewe, M. & Bhardwaj, N. A monocyte conditioned medium is more effective than defined cytokines in mediating the terminal maturation of human dendritic cells. *Blood* **90**, 3640–3646 (1997).
10. Maraskovsky, E. et al. Dramatic increase in the numbers of functionally mature dendritic cells in Flt3

- and ligand-treated mice: Multiple dendritic cell subpopulations identified. *J. Exp. Med.* **184**, 1953–1962 (1996).
11. Adema, G. J. et al. A dendritic-cell-derived C-C chemokine that preferentially attracts naive T cells. *Nature* **387**, 713–717 (1997).
12. Mueller, C. G. et al. Polymerase chain reaction selects a novel disintegrin-proteinase from CD40-activated germinal center dendritic cells. *J. Exp. Med.* **186**, 655–663 (1997).
13. Greaves, D. R. et al. CCR6, a CC chemokine receptor that interacts with macrophage inflammatory protein 3 α and is highly expressed in human dendritic cells. *J. Exp. Med.* **186**, 837–844 (1997).
14. Winzler, C. et al. Maturation stages of mouse dendritic cells in growth factor-dependent long-term cultures. *J. Exp. Med.* **185**, 317–328 (1997).
15. Bhardwaj, N., Young, J. W., Nisanian, A. J., Baggers, J. & Steinman, R. M. Small amounts of superantigen, when presented on dendritic cells, are sufficient to initiate T cell responses. *J. Exp. Med.* **178**, 633–642 (1993).
16. Inaba, K., Inaba, M., Naito, M. & Steinman, R. M. Dendritic cell progenitors phagocytose particulates, including *Bacillus Calmette-Guerin* organisms, and sensitize mice to mycobacterial antigens *in vivo*. *J. Exp. Med.* **178**, 479–488 (1993).
17. Moll, H., Fuchs, H., Blank, C. & Rollinghoff, M. Langerhans cells transport *Leishmania major* from the infected skin to the draining lymph node for presentation to antigen-specific T cells. *Eur. J. Immunol.* **23**, 1595–1601 (1993).
18. Zitvogel, L. et al. Therapy of murine tumors with tumor peptide pulsed dendritic cells: Dependence on T-cells, B7 costimulation, and Th1-associated cytokines. *J. Exp. Med.* **183**, 87–97 (1996).
19. Paglia, P., Chiodoni, C., Rodolfo, M. & Colombo, M. P. Murine dendritic cells loaded *in vitro* with soluble protein prime CTL against tumor antigen *in vivo*. *J. Exp. Med.* **183**, 317–322 (1996).
20. Mayordomo, J. I. et al. Bone marrow-derived dendritic cells pulsed with synthetic tumour peptides elicit protective and therapeutic antitumour immunity. *Nature Med.* **1**, 1297–1302 (1995).
21. Hsu, F. J. et al. Vaccination of patients with B-cell lymphoma using autologous antigen-pulsed dendritic cells. *Nature Med.* **2**, 52–58 (1996).
22. Ingulli, E., Mondino, A., Khoruts, A. & Jenkins, M. K. *In vivo* detection of dendritic cell antigen presentation to CD4⁺ T cells. *J. Exp. Med.* **185**, 2133–2141 (1997).
23. Luther, S. A., Gulbranson-Judge, A., Acha-Orbea, H. & MacLennan, I. C. M. Viral superantigen drives extrafollicular and follicular B differentiation leading to virus-specific antibody production. *J. Exp. Med.* **185**, 551–562 (1997).
24. Kudo, S., Matsuno, K., Ezaki, T. & Ogawa, M. A novel migration pathway for rat dendritic cells from the blood: Hepatic sinusoids-lymph translocation. *J. Exp. Med.* **185**, 777–784 (1997).
25. Inaba, K. et al. High levels of a major histocompatibility complex II-self peptide complex on dendritic cells from lymph node. *J. Exp. Med.* **186**, 665–672 (1997).
26. Cella, M. et al. Ligation of CD40 on dendritic cells triggers production of high levels of interleukin-12 and enhances T cell stimulatory capacity: T-T help via APC activation. *J. Exp. Med.* **184**, 747–752 (1996).
27. Koch, F. et al. High level IL-12 production by murine dendritic cells: upregulation via MHC class II and CD40 molecules and downregulation by IL-4 and IL-10. *J. Exp. Med.* **184**, 741–747 (1996).
28. Reis e Sousa, C. et al. *In vivo* microbial stimulation induces rapid CD40L-independent production of IL-12 by dendritic cells and their re-distribution to T cell areas. *J. Exp. Med.* **186**, 1819–1829 (1997).
29. Caux, C. et al. B70/B7-2 is identical to CD86 and is the major functional ligand for CD28 expressed on human dendritic cells. *J. Exp. Med.* **180**, 1841–1847 (1994).
30. Inaba, K. et al. The tissue distribution of the B7-2 costimulator in mice: abundant expression on dendritic cells *in situ* and during maturation *in vitro*. *J. Exp. Med.* **180**, 1849–1860 (1994).
31. Bhardwaj, N. et al. Influenza virus-infected dendritic cells stimulate strong proliferative and cytolytic responses from human CD8⁺ T cells. *J. Clin. Invest.* **94**, 797–807 (1994).
32. Bender, A., Bui, L. K., Feldman, M. A. V., Larsson, M. & Bhardwaj, N. Inactivated influenza virus, when presented on dendritic cells, elicits human CD8⁺ cytolytic T cell responses. *J. Exp. Med.* **182**, 1663–1671 (1995).
33. Caux, C. et al. Activation of human dendritic cells through CD40 cross-linking. *J. Exp. Med.* **180**, 1263–1272 (1994).
34. Wong, B. R. et al. TRANCE, a new TNF family member predominantly expressed in T cells, is a dendritic cell specific survival factor. *J. Exp. Med.* **186**, 2075–2080 (1997).
35. Anderson, D. M. et al. A homologue of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. *Nature* **390**, 175–179 (1997).
36. Lukas, M. et al. Human cutaneous dendritic cells migrate through dermal lymphatic vessels in a skin organ culture model. *J. Invest. Dermatol.* **106**, 1293–1299 (1996).
37. Reis e Sousa, C., Stahl, P. D. & Austyn, J. M. Phagocytosis of antigens by Langerhans cells *in vitro*. *J. Exp. Med.* **178**, 509–519 (1993).
38. Svensson, M., Stockinger, B. & Wick, M. J. Bone marrow-derived dendritic cells can process bacteria for MHC-I and MHC-II presentation to T cells. *J. Immunol.* **158**, 4229–4236 (1997).
39. Jiang, W. et al. The receptor DEC-205 expressed by dendritic cells and thymic epithelial cells is involved in antigen processing. *Nature* **375**, 151–155 (1995).
40. Nijman, H. W. et al. Antigen capture and MHC class II compartments of freshly isolated and cultured human blood dendritic cells. *J. Exp. Med.* **182**, 163–174 (1995).
41. Pierre, P. et al. Developmental regulation of MHC class II transport in mouse dendritic cells. *Nature* **388**, 787–792 (1997).
42. Cella, M., Engering, A., Pinet, V., Pieters, J. & Lanzavecchia, A. Inflammatory stimuli induce accumulation of MHC class II complexes on dendritic cells. *Nature* **388**, 782–787 (1997).
43. Albert, M. L., Sauter, B. & Bhardwaj, N. Dendritic cells acquire antigen from apoptotic cells and induce class-I restricted CTLs. *Nature* **392**, 86–89 (1998).
44. Buelens, C. et al. Human dendritic cell responses to lipopolysaccharide and CD40 ligation are differentially regulated by IL-10. *Eur. J. Immunol.* **27**, 1848–1852 (1997).
45. Sallusto, F., Nicoletti, C., De Maria, R., Corinti, S. & Testi, R. Ceramide inhibits antigen uptake and presentation by dendritic cells. *J. Exp. Med.* **184**, 2411–2416 (1996).
46. Granelli-Piperno, A., Pope, M., Inaba, K. & Steinman, R. M. Coexpression of REL and SP1 transcription factors in HIV-1 induced, dendritic cell-T cell syncytia. *Proc. Natl. Acad. Sci. USA* **92**, 1094–10948 (1995).
47. Kitajima, T., Arizumi, K., Bergstresser, P. R. & Takashima, A. A novel mechanism of glucocorticoid-induced immune suppression: The inhibition of T cell-mediated terminal maturation of a murine dendritic cell line. *J. Clin. Invest.* **98**, 142–147 (1996).
48. Brocker, T. Survival of mature CD4 T lymphocytes is dependent on MHC class II expressing dendritic cells. *J. Exp. Med.* **186**, 1223–1232 (1997).
49. O'Doherty, U. et al. Human blood contains two subsets of dendritic cells, one immunologically mature, and the other immature. *Immunology* **82**, 487–493 (1994).
50. Brown, K. A. et al. Human blood dendritic cells: binding to vascular endothelium and expression of adhesion molecules. *Clin. Exp. Immunol.* **107**, 601–607 (1997).
51. Matsuno, K., Ezaki, T., Kudo, S. & Uehara, Y. A life stage of particle-laden rat dendritic cells *in vivo*: their terminal division, active phagocytosis and translocation from the liver to hepatic lymph. *J. Exp. Med.* **183**, 1865–1878 (1996).

52. McWilliam, A. S. *et al.* Dendritic cells are recruited into the airway epithelium during the inflammatory response to a broad spectrum of stimuli. *J. Exp. Med.* **184**, 2429–2432 (1996).
53. Roake, J. A. *et al.* Dendritic cell loss from non-lymphoid tissues following systemic administration of lipopolysaccharide, tumour necrosis factor, and interleukin-1. *J. Exp. Med.* **181**, 2237–2248 (1995).
54. MacPherson, G. G., Jenkins, C. D., Stein, M. J. & Edwards, C. Endotoxin-mediated dendritic cell release from the intestine: Characterization of released dendritic cells and TNF dependence. *J. Immunol.* **154**, 1317–1322 (1995).
55. De Smedt, T. *et al.* Regulation of dendritic cell numbers and maturation by lipopolysaccharide *in vivo*. *J. Exp. Med.* **184**, 1413–1424 (1996).
56. Hosoi, J. *et al.* Regulation of Langerhans cell function by nerves containing calcitonin gene-related peptide. *Nature* **363**, 159–162 (1993).
57. Sozzani, S. *et al.* Migration of dendritic cells in response to formyl peptides, C5a, and a distinct set of chemokines. *J. Immunol.* **155**, 3292–3295 (1995).
58. Young, I. W., Szabolcs, P. & Moore, M. A. S. Identification of dendritic cell colony-forming units among normal CD4⁺ bone marrow progenitors that are expanded by *c-kit*-ligand and yield pure dendritic cell colonies in the presence of granulocyte/macrophage colony-stimulating factor and tumor necrosis factor α . *J. Exp. Med.* **182**, 1111–1120 (1995).
59. Saunders, D. *et al.* Dendritic cell development in culture from thymic precursor cells in the absence of granulocyte/macrophage colony-stimulating factor. *J. Exp. Med.* **184**, 2185–2196 (1996).
60. Flores-Romo, L. *et al.* CD40 ligation on human CD34⁺ hematopoietic progenitors induces their proliferation and differentiation into functional dendritic cells. *J. Exp. Med.* **185**, 341–349 (1997).
61. Caux, C. *et al.* CD34⁺ hematopoietic progenitors from human cord blood differentiate along two independent dendritic cell pathways in response to GM-CSF+ TNF α . *J. Exp. Med.* **184**, 695–706 (1996).
62. Strunk, D., Egger, C., Leitner, G., Hanau, D. & Stingl, G. A skin homing molecule defines the Langerhans cells progenitor in human peripheral blood. *J. Exp. Med.* **185**, 1131–1136 (1997).
63. Caux, C. *et al.* CD34⁺ hematopoietic progenitors from human cord blood differentiate along two independent dendritic cell pathways in response to GM-CSF+TNF α : II Functional analysis. *Blood* **90**, 1458–1470 (1997).
64. Szabolcs, P. *et al.* Dendritic cells and macrophages can mature independently from a human bone marrow-derived, post-CFU intermediate. *Blood* **87**, 4520–4530 (1996).
65. Borkowski, T. A., Letterio, J. J., Farr, A. G. & Udey, M. C. A role for endogenous transforming growth factor β 1 in Langerhans cell biology: The skin of transforming growth factor β 1 null mice is devoid of epidermal Langerhans cells. *J. Exp. Med.* **184**, 2417–2422 (1996).
66. Suss, G. & Shortman, K. A subclass of dendritic cells kills CD4 T cells via Fas/Fas-ligand induced apoptosis. *J. Exp. Med.* **183**, 1789–1796 (1996).
67. Ardavin, C., Wu, L., Li, C. & Shortman, K. Thymic dendritic cells and T cells develop simultaneously in the thymus from a common precursor population. *Nature* **362**, 761–763 (1993).
68. Grouard, G. *et al.* The enigmatic plasmacytoid T cells develop into dendritic cells with IL-3 and CD40-ligand. *J. Exp. Med.* **185**, 1101–1111 (1997).
69. Lieberman, P. H. *et al.* Langerhans cell [eosinophilic] granulomatosis. A clinicopathologic study encompassing 50 years. *Am. J. Surg. Pathol.* **20**, 519–552 (1996).
70. Dubois, B. *et al.* Dendritic cells enhance growth and differentiation of CD40-activated B lymphocytes. *J. Exp. Med.* **185**, 941–951 (1997).
71. Fayette, J. *et al.* Human dendritic cells skew isotype switching of CD40-activated naive B cells towards IgA1 and IgA2. *J. Exp. Med.* **185**, 1909–1918 (1997).
72. Frankel, S. S. *et al.* Replication of HIV-1 in dendritic cell-derived syncytia at the mucosal surface of the adenoid. *Science* **272**, 115–117 (1996).
73. Frankel, S. S. *et al.* Active replication of HIV-1 at the lymphoepithelial surface of the tonsil. *Am. J. Pathol.* **151**, 89–96 (1997).
74. Kelsall, B. L. & Strober, W. Distinct populations of dendritic cells are present in the subepithelial dome and T cell regions of the murine Peyer's Patch. *J. Exp. Med.* **183**, 237–247 (1996).
75. Matsumoto, M. *et al.* Distinct roles of lymphotoxin- α and type 1 TNF receptor in the establishment of follicular dendritic cells from non-bone marrow-derived cells. *J. Exp. Med.* **186**, 1997–2004 (1997).
76. Liu, Y.-J., Grouard, G., de Bouteiller, O. & Banchereau, J. Follicular dendritic cells and germinal centers. *Int. Rev. Cytology* **166**, 139–179 (1996).
77. Liu, Y.-J. *et al.* Follicular dendritic cells specifically express the long CR2/CD21 isoform. *J. Exp. Med.* **185**, 165–170 (1997).
78. Grouard, G., Durand, I., Filgueira, L., Banchereau, J. & Liu, Y.-J. Dendritic cells capable of stimulating T cells in germinal centres. *Nature* **384**, 364–367 (1996).
79. Brocker, T. & Karjalainen, K. Targeted expression of MHC class II molecules demonstrates that dendritic cells can induce negative but no positive selection of thymocytes *in vivo*. *J. Exp. Med.* **185**, 541–550 (1997).
80. Laufer, T. M., DeKoning, J., Markowitz, J. S., Lo, D. & Glimcher, L. H. Unopposed positive selection and autoreactivity in mice expressing class II MHC only on thymic cortex. *Nature* **383**, 81–85 (1996).
81. Zal, T., Volkman, A. & Stockinger, B. Mechanisms of tolerance induction in major histocompatibility complex class II-restricted T cells specific for a blood-borne self-antigen. *J. Exp. Med.* **180**, 2089–2099 (1994).
82. Kurts, C. *et al.* Constitutive class I-restricted exogenous presentation of self antigens *in vivo*. *J. Exp. Med.* **184**, 923–930 (1996).
83. Kurts, C., Kosaka, H., Carbone, F. R., Miller, J. F. A. P. & Heath, W. R. Class I-restricted cross-presentation of exogenous self antigens leads to deletion of autoreactive CD8⁺ T cells. *J. Exp. Med.* **186**, 239–245 (1997).
84. Forster, I. & Lieberman, I. Peripheral tolerance of CD4 T cells following local activation in adolescent mice. *Eur. J. Immunol.* **26**, 3194–3202 (1996).
85. Soderberg-Naucler, C., Fish, K. N. & Nelson, J. A. Reactivation of latent human cytomegalovirus by allogeneic stimulation of blood cells from healthy donors. *Cell* **91**, 119–126 (1997).
86. Rettig, M. B. *et al.* Kaposi's sarcoma-associated herpesvirus infection of bone marrow dendritic cells from multiple myeloma patients. *Science* **276**, 1851–1854 (1997).
87. Cameron, P. U. *et al.* Dendritic cells exposed to human immunodeficiency virus type-1 transmit a vigorous cytopathic infection to CD4⁺ T cells. *Science* **257**, 383–387 (1992).
88. Weissman, D., Barker, T. D. & Fauci, A. S. The efficiency of acute infection of CD4⁺ T cells is markedly enhanced in the setting of antigen-specific immune activation. *J. Exp. Med.* **183**, 687–692 (1996).
89. Pinchuk, L. M., Polacino, P. S., Agy, M. B., Klaus, S. J. & Clark, E. A. The role of CD40 and CD80 accessory cell molecules in dendritic cell-dependent HIV-1 infection. *Immunity* **1**, 317–325 (1994).
90. Pope, M. *et al.* Conjugates of dendritic cells and memory T lymphocytes from skin facilitate productive infection with HIV-1. *Cell* **78**, 389–398 (1994).
91. Grosjean, I. *et al.* Measles virus infects human dendritic cells and blocks their allostimulatory properties for CD4⁺ T cells. *J. Exp. Med.* **186**, 801–812 (1997).
92. Schnorr, J.-J. *et al.* Induction of maturation of human blood dendritic cell precursors by measles virus is associated with immunosuppression. *Proc. Natl Acad. Sci. USA* **94**, 5326–5331 (1997).
93. Fugier-Vivier, I. *et al.* Measles virus suppresses cell-mediated immunity by interfering with the survival and functions of dendritic and T cells. *J. Exp. Med.* **186**, 813–823 (1997).
94. Pope, M., Elmore, D., Ho, D. & Marx, P. Dendritic cell–T cell mixtures, isolated from the skin and mucosae of macaques, support the replication of HIV. *AIDS Res. Hum. Retro.* **13**, 819–827 (1997).
95. Chau, P., Moutet, M., Faivre, J., Martin, F. & Martin, M. Inflammatory cells infiltrating human colorectal carcinomas express HLA class II but not B7-1 and B7-2 costimulatory molecules of the T-cell activation. *Lab. Invest.* **74**, 975–983 (1996).
96. Specht, J. M. *et al.* Dendritic cells retrovirally transduced with a model tumor antigen gene are therapeutically effective against established pulmonary metastases. *J. Exp. Med.* **186**, 1213–1221 (1997).
97. Song, W. *et al.* Dendritic cells genetically modified with an adenovirus vector encoding the cDNA for a model tumor antigen induce protective and therapeutic antitumor immunity. *J. Exp. Med.* **186**, 1247–1256 (1997).
98. Schuler, G. & Steinman, R. M. Dendritic cells as adjuvants for immune-mediated resistance to tumors. *J. Exp. Med.* **186**, 1183–1187 (1997).
99. Josien, R., Heslan, M., Soullou, J.-P. & Cuturi, M.-C. Rat spleen dendritic cells express natural killer cell receptor protein 1 (NKR-P1) and have cytotoxic activity to select targets via a Ca²⁺-dependent mechanism. *J. Exp. Med.* **186**, 467–472 (1997).
100. Condon, C., Watkins, S. C., Celluzzi, C. M., Thompson, K. & Falo, L. D. Jr DNA-based immunization by *in vivo* transfection of dendritic cells. *Nature Med.* **2**, 1122–1128 (1996).
101. Casares, S., Inaba, K., Brumeau, T., Steinman, R. M. & Bona, C. A. Antigen presentation by dendritic cells following immunization with DNA encoding a class II-restricted viral epitope. *J. Exp. Med.* **186**, 1481–1486 (1997).
102. Colonna, M. *et al.* A common inhibitory receptor for MHC class I molecules on human lymphoid and myelomonocytic cells. *J. Exp. Med.* **186**, 1809–1818 (1997).
103. Vicari, A. P. *et al.* TECK: A novel CC chemokine specifically expressed by thymic dendritic cells and potentially involved in T cell development. *Immunity* **7**, 291–302 (1997).

Acknowledgements. We dedicate this review to the memory of J. Chiller. J.B. acknowledges the members of the Schering Plough Laboratory for Immunological Research, Dardilly, and particularly F. Briere, C. Caux, S. Lebecque, Y.-J. Liu, M. Vatan, A. Waitz, and R.S. acknowledges the long-standing contributions of N. Bhardwaj, A. Granelli-Piperno, K. Inaba, I. Mellman, C. Moberg, M. Nussenzweig, M. Pack, M. Pope, N. Romani, G. Schuler, J. Young and the support of the NIAID. Because of space limitations, the reference list has been restricted to a select sample from the past five years. A more complete list is available from the authors. We thank our many colleagues for their critical comments on this review.

Clumpy star-forming regions as the origin of the peculiar morphology of high-redshift galaxies

Masafumi Noguchi

Astronomical Institute, Tohoku University, Aoba-ku, Sendai 980-77, Japan

Many high-redshift galaxies have peculiar morphologies and photometric properties^{1–5}. It is not clear whether these peculiarities originate in galaxy–galaxy interactions (or mergers) or are intrinsic to the galaxies, a natural consequence of the star formation process in primeval systems. Here I report the results of numerical simulations of protogalaxy evolution, which show that the gas-rich disk of a young galaxy becomes gravitationally unstable and fragments into massive clumps of sub-galactic size. Most of the stars are formed in these discrete clumps, thereby providing a natural explanation for the peculiar morphology of high-redshift galaxies. The dynamical evolution of these young systems is dominated by the clumps and ultimately leads to structures resembling present-day galaxies, with a spheroidal bulge and an exponential disk. I interpret the differences between the Hubble types of galaxies as resulting from different timescales of disk formation. Finally, the model provides a causal link between the emergence of quasar activity and the dynamical evolution of the host galaxy.

The protogalaxy was modelled as an ensemble of numerous gas clouds and dark-matter particles distributed uniformly in a spherical volume. The initial radius of the galaxy is 15 kpc and the total mass of the galaxy is $1.5 \times 10^{11} M_{\odot}$, where $M_{\odot}$ is the solar mass. A uniform (that is, solid-body) rotation with an angular frequency $\Omega = 13.9 \text{ km s}^{-1} \text{ kpc}^{-1}$ was given so as to bring the system into nearly centrifugal equilibrium. Inelastic collisions were introduced between gas clouds to simulate the dissipative nature of the gas, whereas the dark-matter particles are assumed to move in a collisionless manner in the galaxy gravitational field. Every pair of two overlapping gas clouds are made to collide with the same restitution coefficient of 0.01, though the outcome of a cloud–cloud collision may in fact be more complicated, depending on the impact parameter and relative velocity⁶. The star formation process is incorporated into this simulation by converting each gas cloud into a stellar particle with a probability proportional to $m_g(\rho_g)^{1/2}$, where m_g is the mass of that cloud and ρ_g is the smoothed-out gas density in its vicinity. A threshold gas density of $0.1 M_{\odot} \text{ pc}^{-3}$ was specified, below which no star formation is allowed. Stellar particles are treated as collisionless. Star formation is expected to affect the kinematics of the interstellar medium through the energy input from stellar winds and supernova explosions. This process is mimicked by making a newly born stellar particle give velocity boosts of $\sim 10 \text{ km s}^{-1}$ to nearby gas clouds. This value of the boost is the one which balances energy dissipation by cloud–cloud collisions in the numerical model for a present-day spiral galaxy⁷. Initially, the masses of the gas component and the dark halo are identical, and there is no star.

As seen in Fig. 1, the galaxy collapses nearly perpendicularly to the galactic plane within $\sim 1 \text{ Gyr}$, and produces a rotating disk. The most remarkable feature of this model is that some ten distinctive clumps are formed in the disk as the disk is built up by the gas infall to the galactic plane. The typical mass of individual clumps measured from the simulation is $10^9 M_{\odot}$. This is larger by three orders of magnitude than the masses of globular clusters and giant molecular clouds, which are the most massive entities known in

present-day galaxies, and is comparable with the masses of dwarf galaxies. These sub-galactic clumps are created by the local gravitational instability of the mostly gaseous galactic disk, which can cool rapidly due to efficient energy dissipation. The existence of such gas-rich galactic disks at early cosmological epochs seems to be vindicated by the recent finding^{8,9} of damped Lyman- α systems at redshifts $z \approx 2\text{--}4$, if the interpretation is correct that these absorption systems represent embryonic galactic disks located on the lines of sight toward more distant quasars. The star formation rate reaches a maximum of $\sim 40 M_{\odot} \text{ yr}^{-1}$ at the main epoch of disk formation ($t \sim 10$) and declines afterwards (Fig. 2).

The clumpy and sometimes asymmetric disk seen in this model is strongly reminiscent of the images of many galaxies at large redshift obtained by the Hubble Space Telescope and ground-based telescopes^{1–5,10}. Steidel *et al.*¹¹ investigated the morphology of Lyman-break galaxies with $2.3 < z < 3.4$ found in the Hubble Deep Field and noticed objects having multicomponent structure. Spectroscopic detection of a systematic rotational motion of clumps in each object would lend a strong support to the interpretation that the peculiar morphology of high-redshift galaxies is a manifestation of gravitational instability in young galactic disks.

These sub-galactic clumps play an important role in the evolution of disk galaxies as described below. The important point here is the longevity of these massive clumps. Though the feedback from internal star formation is considered to destroy the mother molecular clouds in nearby galaxies, the massive clumps formed in early evolutionary phases are likely to survive due to their large binding energy, as the simulation suggests. Although the clumps rotate around the galactic centre in the disk plane, they sometimes merge with each other and evolve into more massive clumps.

Due to their large masses, the clumps experience strong dynamical friction against ambient stars and gas clouds, and spiral towards the galactic centre. Accumulation of clumps at the galactic centre makes a spheroidal bulge (Fig. 1, $t = 29$). Figure 3 shows the distribution of the age for the bulge stars at several epochs. At the final epoch indicated, $t = 28$, the age spread is as large as $\sim 1 \text{ Gyr}$. The bulge is not formed quickly in a single event as the protogalaxy collapses, but is assembled gradually and stepwise as individual clumps orbiting in the disk plane accrete. The nature and evolution of bulges remain one of the most important unresolved problems in galactic astronomy. Recent observations suggest that the bulge of the Milky Way galaxy has a large age range and shows a correlation between abundance and kinematics of the constituent stars¹²; this is difficult to explain by the standard picture¹³ in which the bulge formation is a dynamically rapid process associated with the initial collapse of the protogalaxy. The clump-driven evolution model presented here may explain the enigmatic nature of the Milky Way bulge.

The clumpy disk has a tendency to acquire an exponential surface density distribution (Fig. 4), which is a universal feature of disk galaxies¹⁴. This may be understood, because the dynamical effects of the massive clumps can be viewed as the action of ‘viscosity’^{15,16}, and the viscous process coupled with star formation leads to exponential density profiles in the stellar disks under a variety of initial conditions¹⁷. As the bulge grows, the rotation curve of the disk component gradually steepens and the radial zone having a constant rotational velocity widens, leading to a flat rotation curve, which is commonly observed in disk galaxies, especially of early morphological types¹⁸.

Although the clumps are orbiting in the disk plane, stars and gas clouds are scattered off them and acquire random motions because of the large masses of individual clumps. The old stellar disk component thus shows a progressively large scale height as time elapses. This thickening is a back reaction of the dynamical friction. At the epoch $t = 30$, the scale height of the old stellar group is $\sim 1.9 \text{ kpc}$, which is comparable with that of the thick disk component found in some nearby galaxies^{19,20}.

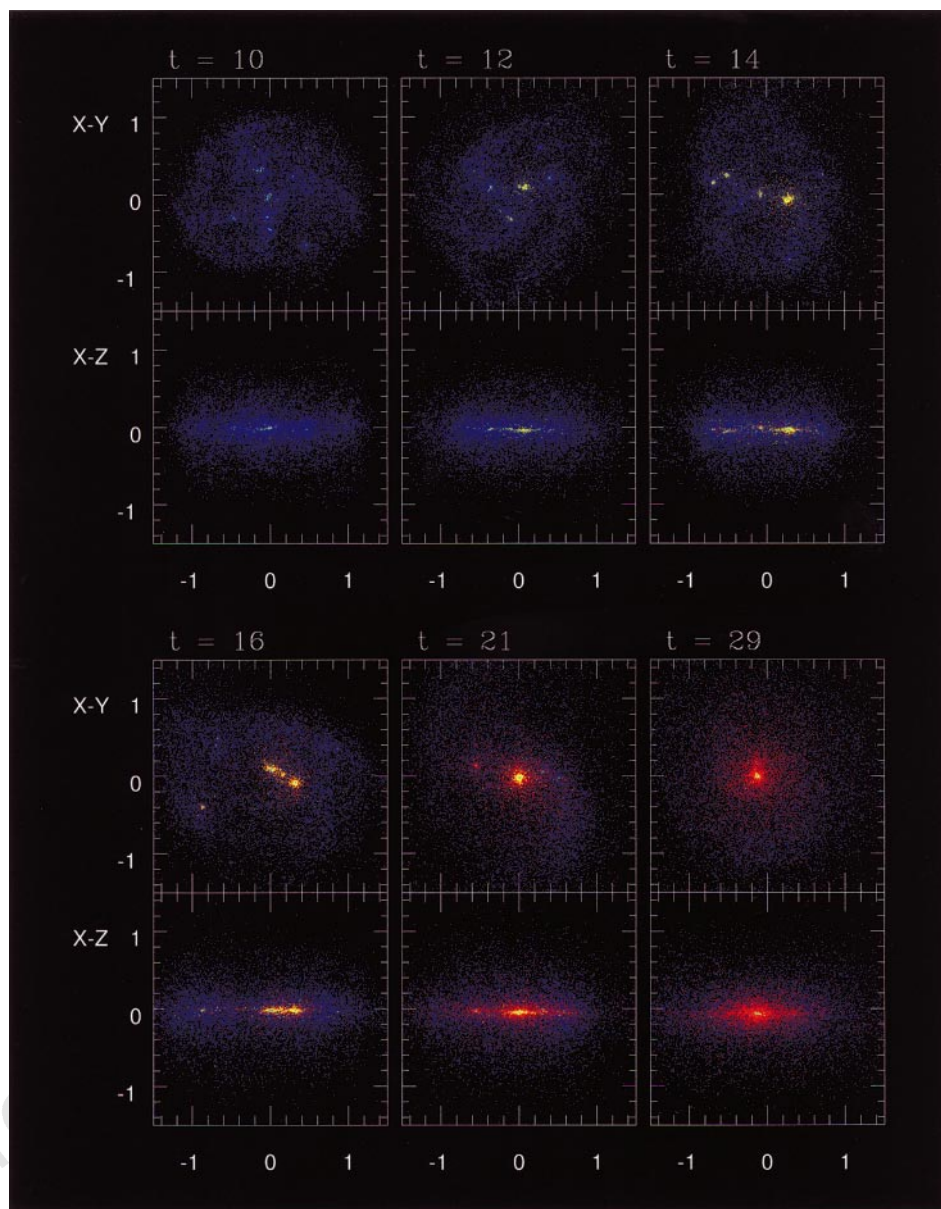


Figure 1 Morphological evolution of a young galaxy model. The gas cloud particles and the stellar particles are projected onto the x - y plane (that is, the disk plane) and the x - z plane. Time (t) and positions are given in the non-dimensional unit system, in which the mass and initial radius of the galaxy are unity. The gravitational constant is unity. One time unit is $\sim 7 \times 10^7$ yr if scaled to a typical galaxy. Gas clouds are represented by blue dots. Stellar particles are colour-coded according to their age, with cyan (greenish-blue) dots representing stars younger than 10^8 yr, yellow dots stars younger than 5×10^8 yr, and red dots other

older stars. The initial system is composed of 30,000 gas clouds and 20,000 dark-matter particles. Each cloud has a mass of $2.5 \times 10^6 M_\odot$ and a radius of 38 pc, being typical of giant molecular clouds. Collisions between two gas clouds are assumed to be highly inelastic, with the restitution coefficient $f_{\text{col}} = 0.01$ in this model. Evolution of the model remains essentially the same for the range, $f_{\text{col}} < 0.5$, whereas a larger coefficient leads to slower formation and less clumpy structure of the disk component.

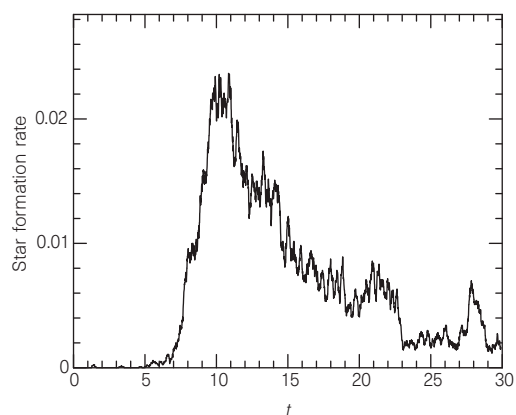


Figure 2 Time variation of star formation rate. Time, t , and star formation rate are given in non-dimensional units. For a typical disk galaxy of a mass $1.5 \times 10^{11} M_\odot$ and a radius 15 kpc, unit time is $\sim 7 \times 10^7$ yr and the star formation rate of 0.01 is $\sim 20 M_\odot \text{ yr}^{-1}$. Two moderate starbursts at $t \sim 21$ and ~ 28 are caused by infall of a sub-galactic clump into the bulge.

The clump-driven evolution model provides a unique prediction of the emergence of active galactic nuclei (AGN) in evolving galaxies. The standard picture of AGN (including quasars) presumes, as their engines, a super-massive black hole located at the galactic centre and fed by gas inflow from the surrounding accretion disk at a rate of $\sim 1 M_{\odot} \text{ yr}^{-1}$ (ref. 21). The mergers and accumulation of the heavy sub-galactic clumps may provide an efficient way to cause the required inflow. Figure 5 indeed shows that each merger event induces accretion of gas onto the merging pair with a sufficiently high inflow rate. These massive clumps undergoing mergers can be regarded as potential sites of quasar activity. In this case, the quasar activity is associated closely with the clump mergers and formation of the bulge, and a wide variety is expected in the emergence of activity as follows. The active site need not coincide with the galaxy nucleus defined by the centre of the bulge component, because a clump merger can happen in the outskirts of the galactic disk. In principle, there is even a possibility of multiple quasars. For example, a double quasar may be formed when two pairs of colliding clumps happen to merge at nearly the same instant. The galactic nucleus itself may be activated when one of the orbiting clumps has fallen into the bulge region. Thus the growth of the black holes within a galaxy may be hierarchical, from numerous small seed black holes which are embedded in clumps and orbiting in the disk, to successively larger ones via mergers, and

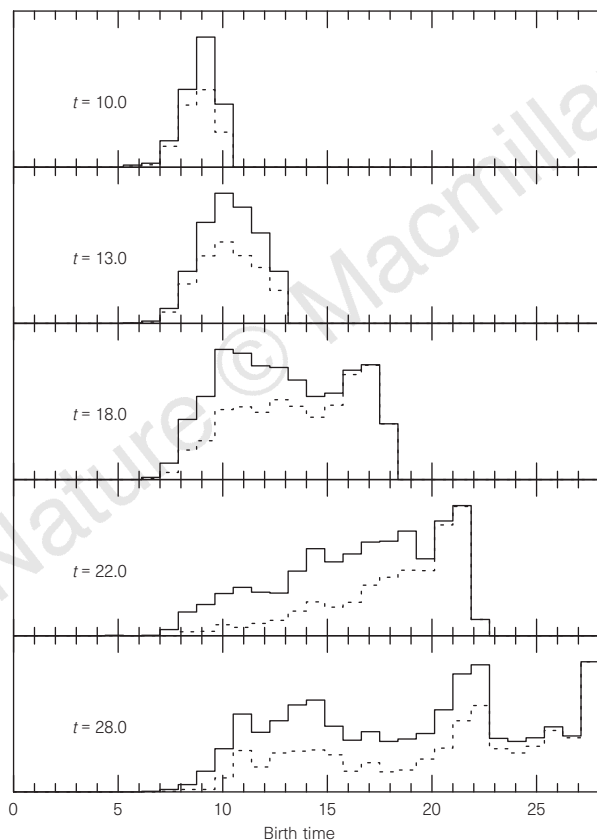


Figure 3 Distribution of the age for the stars contained in the bulge region. Each histogram shows, for the given epoch, the relative number of stars which were born at the time indicated on the abscissa. Solid lines are distributions for stars located within 0.1 length units of the bulge centre, whereas the dotted lines are for stars located within 0.025. Times are given in non-dimensional units. Note that the bulge sometimes contains a significant amount of young stars with ages $< 10^8$ yr. This is due to the starburst caused by the falling of one of the massive clumps into the bulge region. Signature of each accretion is seen as a local maximum in the age distribution. Formation of the bulge through successive mergers of sub-galactic clumps may explain the blue nucleated galaxies, which Schade²⁷ noted among high-redshift galaxy samples and which have a blue compact central region surrounded by a redder envelope.

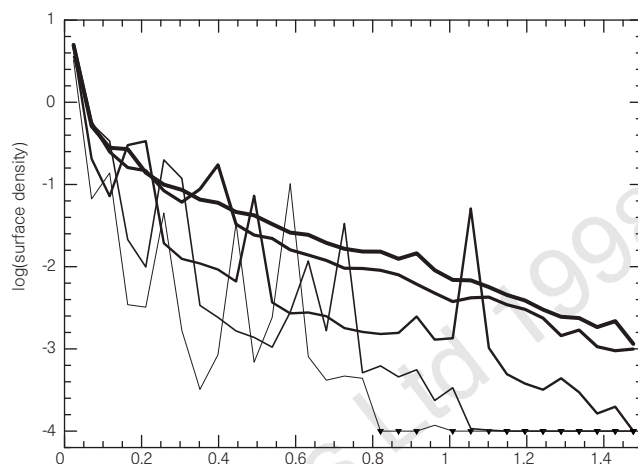


Figure 4 Growth of the disk component having an exponentially decreasing surface density profile. The surface density distribution in the projection onto the x - y plane for the stars is plotted as a function of the galactocentric distance r . The progressively thicker lines correspond to the epochs $t = 11, 13, 15, 22$ and 29 . For a typical disk galaxy of a mass $1.5 \times 10^{11} M_{\odot}$ and a radius 15 kpc, unit time is $\sim 7 \times 10^7$ yr. Radius and densities are given in the non-dimensional unit system. Filled triangles indicate upper limits.

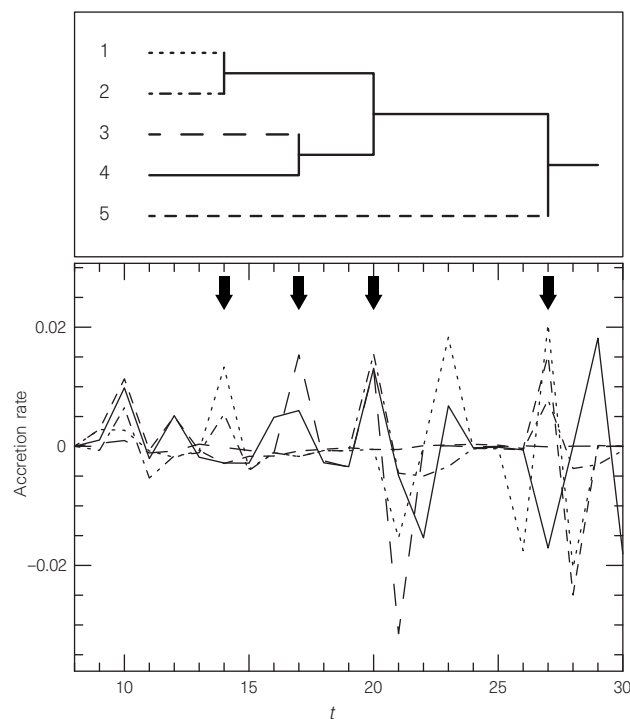


Figure 5 Merger history of the five prominent clumps identified at $t = 9$ (top panel) and the mass accretion onto individual clumps (bottom panel). The same clump is indicated by the same line type in the upper and lower panels. Significant accretion events associated with mergers between two clumps are indicated by arrows. The maximum accretion rate is as large as 0.01, which corresponds to $\sim 20 M_{\odot} \text{ yr}^{-1}$ for a typical disk galaxy of a mass $1.5 \times 10^{11} M_{\odot}$ and a radius 15 kpc (unit time is $\sim 7 \times 10^7$ yr). Each merging pair could provide a site of quasar activity. In this view, there is no single well-defined epoch of quasar activity, and one galaxy can be activated many times, repeatedly and intermittently.

finally to one (or a few) super-massive black hole(s) located in the nucleus.

What is the meaning of the Hubble morphological sequence in the framework of the clump-driven evolution model presented here? I have proposed²² that one of key parameters which determine the morphological type of the resulting galaxy is the timescale of disk formation (although I must note that disk galaxies do not constitute a one-parameter family²³). The masses of gas clumps become smaller if the disk formation timescale is increased because the disk contains less gas at the main formation epoch (giant molecular associations such as found in M51 may correspond to these less massive clumps). Thus all the clump-driven dynamical processes are subdued. Moreover, smaller clumps are considered to be more vulnerable to dispersal by star formation activity. A slow-collapse model thus develops a smaller bulge component, has a thinner stellar disk and a rotation curve rising outwards more gently, and ends with a larger gas fraction at the present epoch. These features are all characteristic of late-type disk galaxies.

The present model predicts positive correlations between the relative bulge dominance, existence of a thick disk, and the mass of the super-massive black hole at the galactic centre. Observational studies^{20,24,25} have found marginal evidence that thick disk components are more common in disk galaxies having an appreciable bulge. Today, the mass of the black hole at the galactic centre is being measured in an increasingly larger number of galaxies, using various methods. Though the number of sample galaxies is still small, there is indication of a positive correlation between the black hole mass and the mass of the host bulge²⁶.

The galaxy evolution models constructed by Larson¹³ have long served as a theoretical framework in extragalactic astronomy. In his models, the early efficient star formation produces a bulge as the galaxy collapses; the disk is formed from the residual gas which is allowed to condense into the galactic plane as the star formation process is slowed down for some unknown reason. Larson needed to assume two separate regimes of star formation to explain the two-component nature of disk galaxies: one important feature of the clumpy galaxy model I present here is that it can explain the global properties of galaxies at both high and low redshifts, without resorting to unusual regimes of star formation. □

Received 25 March 1997; accepted 5 January 1998.

1. Abraham, R. G. *et al.* Galaxy morphology to $I = 25$ mag in the Hubble Deep Field. *Mon. Not. R. Astron. Soc.* **279**, L47–L52 (1996).
2. Cowie, L. L., Hu, E. M. & Songaila, A. Faintest galaxy morphologies from HST WFPC2 imaging of the Hawaii Survey Fields. *Astron. J.* **110**, 1576–1583 (1995).
3. Franx, M., Illingworth, G. D., Kelson, D. D., van Dokkum, P. G. & Tran, K.-V. A pair of lensed galaxies at $z = 4.92$ in the field of CL 1358+62. *Astrophys. J.* **486**, L75–L77 (1997).
4. Griffiths, R. E. *et al.* The morphology of faint galaxies in Medium Deep Survey images using WFPC2. *Astrophys. J.* **435**, L19–L22 (1994).
5. van den Bergh, S. *et al.* A morphological catalog of galaxies in the Hubble Deep Field. *Astron. J.* **112**, 359–368 (1996).
6. Lattanzio, J. C., Monaghan, J. J., Pongracic, H. & Schwarz, M. P. Interstellar cloud collisions. *Mon. Not. R. Astron. Soc.* **215**, 125–147 (1985).
7. Noguchi, M. & Ishibashi, S. Simulations of close encounter between galaxies: behaviour of interstellar gas clouds and enhancement of star formation rate. *Mon. Not. R. Astron. Soc.* **219**, 305–331 (1986).
8. Lu, L., Sargent, W. L. W., Womble, D. S. & Barlow, T. A. Properties of a high-redshift galaxy at $z = 4.4$. *Astrophys. J.* **457**, L1–L4 (1996).
9. Wolfe, A. M. *et al.* Metal abundances and kinematics of a high-redshift galaxy obtained with the Keck telescope. *Astrophys. J.* **435**, L101–L104 (1994).
10. Pascarelle, S. M., Windhorst, R. A., Keel, W. C. & Odewahn, S. C. Sub-galactic clumps at a redshift of 2.39 and implications for galaxy formation. *Nature* **383**, 45–50 (1996).
11. Steidel, C. C., Giallisco, M., Dickinson, M. & Adelberger, K. L. Spectroscopy of Lyman break galaxies in the Hubble Deep Field. *Astron. J.* **112**, 352–358 (1996).
12. Rich, R. M. in *Unsolved Problems of the Milky Way* (eds Blitz, L. & Teuben, P.) 403–410 (IAU Symp. No. 169, Kluwer Academic, Dordrecht, 1996).
13. Larson, R. B. Models for the formation of disc galaxies. *Mon. Not. R. Astron. Soc.* **176**, 31–52 (1976).
14. Freeman, K. C. On the disks of spiral and S0 galaxies. *Astrophys. J.* **160**, 811–830 (1970).
15. Fukunaga, M. Viscous flow in cold self-gravitating gas disk. *Publ. Astron. Soc. Jpn* **41**, 965–974 (1989).
16. Lin, D. N. C. & Pringle, J. E. A viscosity prescription for a self-gravitating accretion disc. *Mon. Not. R. Astron. Soc.* **225**, 607–613 (1987).
17. Saio, H. & Yoshii, Y. Viscous evolution of self-gravitating galactic disks within a dark halo. *Astrophys. J.* **363**, 40–49 (1990).
18. Rubin, V. C., Burstein, D., Ford, W. K. & Thonnard, N. Rotation velocities of 16 Sa galaxies and a comparison of Sa, Sb, and Sc rotation properties. *Astrophys. J.* **289**, 81–104 (1985).
19. Burstein, D. Structure and origin of S0 galaxies. III. The luminosity distribution perpendicular to the plane of the disks in S0's. *Astrophys. J.* **234**, 829–836 (1979).
20. Tsikoudi, V. Photometry and structure of lenticular galaxies. I. NGC 3115. *Astrophys. J.* **234**, 842–853 (1979).

21. Blandford, R. D. & Rees, M. J. in *Testing the AGN Paradigm* (eds Holt, S. S., Neff, S. G. & Urry, C. M.) 3–19 (Am. Inst. Phys., New York, 1992).
22. Noguchi, M. Barred galaxies: Intrinsic or extrinsic? *Astrophys. J.* **469**, 605–622 (1996).
23. van den Bergh, S. A preliminary luminosity classification for galaxies of type Sb. *Astrophys. J.* **131**, 558–573 (1960).
24. van der Kruit, P. C. & Searle, L. Surface photometry of edge-on spiral galaxies. I. A model for the three-dimensional distribution of light in galactic disks. *Astron. Astrophys.* **95**, 105–115 (1981).
25. van der Kruit, P. C. & Searle, L. Surface photometry of edge-on spiral galaxies. II. The distribution of light and color in the disk and spheroid of NGC 891. *Astron. Astrophys.* **95**, 116–126 (1981).
26. Kormendy, J. & Richstone, D. Inward bound — The search for supermassive black holes in galactic nuclei. *Ann. Rev. Astron. Astrophys.* **33**, 581–624 (1995).
27. Schade, D. *et al.* Canada-France Redshift Survey: Hubble Space Telescope imaging of high-redshift field galaxies. *Astrophys. J.* **451**, L1–L4 (1995).

Acknowledgements. This work was supported in part by a Grant-in-Aid for Scientific Research from The Ministry of Education, Science, Sports and Culture of Japan.

Correspondence and requests for materials should be addressed to M.N. (e-mail address: noguchi@astroa.astro.hokudai.ac.jp).

Evidence against turbulent and canopy-like magnetic fields in the solar chromosphere

Egidio Landi Degl'Innocenti

Dipartimento di Astronomia e Scienza dello Spazio, Università di Firenze, Italia

Measurements of the degree of polarization of light in the spectral lines emitted by gas near the Sun's limb (its outer edge) can be used to investigate the electron densities and magnetic field strengths in the solar atmosphere; these quantities are important for determining the balance and transport of energy through the Sun's atmosphere. Recent measurements^{1,2} revealed a surprising degree of polarization in the sodium doublet; these observations have remained an enigma. Here I report a mechanism that may explain these observations, in which it is assumed that the populations of the magnetic sublevels of the electronic ground state of the sodium atom are not equal: this leads to ground-level atomic polarization. This mechanism explains very well the observed line shapes, and implies that depolarization does not occur in the solar chromosphere, which would seem to rule out the existence of turbulent magnetic fields and of horizontal, canopy-like fields stronger than ~ 0.01 G. This is difficult to understand, because there is substantial evidence from other types of observation for both types of field^{3–10}. There are obviously aspects of the Sun's atmosphere that remain very poorly understood.

Radiation from the solar limb is linearly polarized, the electric field being perpendicular to the solar radius. The fractional polarization is small, typically ranging from 10^{-4} in the continuum to 10^{-2} in resonance lines. The standard scheme used to interpret line polarization is based on the two-level-atom approximation and on the assumption that all the magnetic sublevels of the ground level have the same population (that is, absence of atomic polarization in the ground level)⁴. The anisotropy of the radiation field, acting on such a model atom, induces differences in the populations of the magnetic sublevels of the upper level and well defined phase relationships, or coherences, between them (that is, upper-level atomic polarization). When the atom de-excites, a characteristic polarization profile results; for the case of the Na I doublet, such a profile is shown in Fig. 1a (solid line). This profile depends on a few parameters: one is the anisotropy factor of the radiation field, w , assumed here independent of wavelength and of the depth in the atmosphere; another parameter is the amount of slightly polarized continuum emission that adds to the line emission of the sodium doublet. A third parameter is the Doppler width, $\Delta\lambda_D$, due both to thermal and turbulent motions, which produces a typical flattening in the cores of the D_1 and D_2 lines.

Recent observations^{1,2} of the limb of the quiet Sun, performed with the high-sensitivity polarimeter ZIMPOL (Zürich imaging polarimeter), show however that the polarization profile of the sodium doublet (Fig. 1b, solid line) differs considerably from the one just presented (Fig. 1a, solid line). Interpretation of these ZIMPOL observations requires the taking into account of two different effects: hyperfine structure and ground-level atomic polarization, a phenomenon neglected in the standard scheme. When hyperfine structure is accounted for—still without atomic polarization in the ground level—the polarization profile is changed to the dashed line of Fig. 1a, which shows the depolarizing effect of hyperfine structure on the D₂ line. To interpret the existence of sharp polarization peaks in the cores of the D lines, we have on the contrary to resort to the second effect, namely the presence of atomic polarization in the ground level. Indeed, such a phenomenon is well known in laboratory spectroscopy¹¹ where the physical mechanism responsible for its appearance is usually referred to as ‘depolarization pumping’. In the absence of depolarizing collisions, the anisotropy of the radiation field is indeed capable of introducing in the ground level atomic polarization of the same order of magnitude as in the excited level, as shown by simple order-of-magnitude estimates¹² and by detailed radiative transfer calculations performed for a two-level model atom¹³.

There are two different mechanisms by which atomic polariza-

tion in the hyperfine structure components of the ground level could affect the polarization in the sodium doublet. The first is the transfer of atomic polarization from the ground level to the excited level via the isotropic component of the radiation field. This brings a contribution to the atomic polarization of the excited level that adds to that considered in the standard scheme and leads to a modification of the polarization of the radiation emitted by the atom. The second mechanism is a radiation transfer effect: the fact that the magnetic sublevels of the ground level are unequally populated introduces dichroism in the solar atmosphere, and thus contributes to the modification of the emergent profile. Both mechanisms are capable of leading to sharp polarization peaks in the cores of the D₁ and D₂ lines.

Here I do not try to solve this complicated spectroscopic problem in detail: rather, I consider a simple model, depending on a few parameters, which is capable of accounting for all the physical mechanisms involved. I first observe that, owing to the cylindrical symmetry of the problem (in the quiet solar atmosphere, the vertical to the surface is a symmetry axis) and to the weak anisotropy of the radiation field, the atomic polarization of the ground level can be simply described in terms of ‘alignment’ coefficients. Such coefficients are proportional to the anisotropy factor w of the radiation field. The alignment coefficients are given by the ratio ρ_0^2/ρ_0^0 of the statistical tensor of the density matrix and can be simply related to the populations of the different magnetic sublevels. I introduce for this purpose two parameters, β_1 and β_2 , which respectively describe the alignment coefficients of the $F = 1$ and $F = 2$ hyperfine sublevels of the ground level:

$$\left[\frac{\rho_0^2}{\rho_0^0} \right]_{F=1} = \beta_1 w, \quad \left[\frac{\rho_0^2}{\rho_0^0} \right]_{F=2} = \beta_2 w \quad (1)$$

The two parameters β_i will be adjusted to find a best fit to the observations and no attempt will be made to check whether their values are consistent with the anisotropy of the radiation field. This would require considering the statistical equilibrium equation connecting the lower-level density matrix to the radiation field, which implies complicated radiative transfer calculations that are outside the scope of this Letter.

Once the ground-level populations are parametrized through equations (1), I proceed, according to the theory presented in ref. 14, to find the dichroism coefficients appearing in the transfer equation for polarized radiation, and, after having solved the statistical equilibrium for the upper level, to find the emissivity coefficients appearing in the same equation. Supposing that the sublevels of the lower level are infinitely sharp, that there are no depolarizing collisions, and that β_1 and β_2 do not depend on depth, the transfer equation can be solved and the fractional linear polarization resulting in the sodium doublet can be expressed, to first order in w , in the following form:

$$Q/I = w(x_v + x'_v - x''_v) \quad (2)$$

Here Q and I are the Stokes parameters defined as in ref. 15, x_v is the emissivity in Stokes Q (normalized to the emissivity in Stokes I) resulting from the standard scheme, x'_v is the analogous quantity caused by the presence of atomic polarization in the ground level, and x''_v is the absorption coefficient in Stokes Q (normalized to the absorption coefficient in Stokes I) resulting again from the atomic polarization of the ground level (dichroism). The three quantities x_v , x'_v and x''_v are complicated expressions involving several Wigner coefficients (x_v and x'_v , for instance, contain one 3-j symbol, nine 6-j symbols, and one 9-j symbol each). They can be computed through the formalism described in ref. 14. Moreover, x'_v and x''_v are linear combinations of the parameters β_i and vanish when the ground-level atomic polarization is zero.

The assumption that the parameters β_i are independent of depth, implicit in equation (2), is however too crude. A better approximation is obtained by taking into account the effect of depolarizing

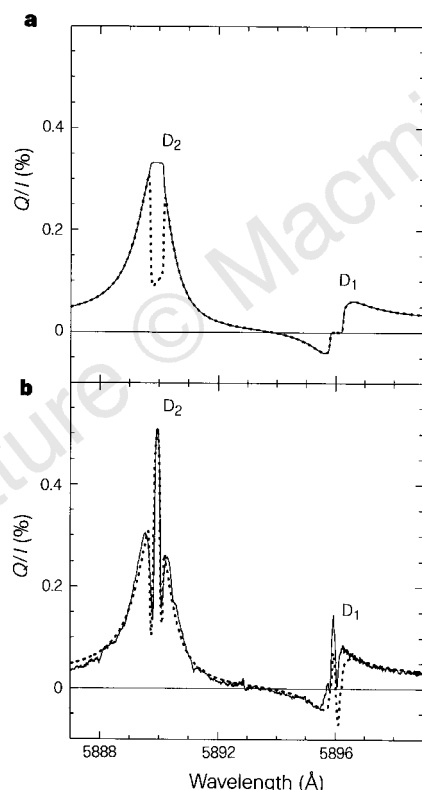


Figure 1 Polarization profiles Q/I , with the positive Q -direction defined as the tangent to the solar limb. In **a**, the solid line is the profile resulting from the standard interpretative scheme that neglects atomic polarization in the lower level (see text); the dashed line takes into account hyperfine structure in the same scheme. The data are obtained for $w = 6.25 \times 10^{-3}$, $\Delta\lambda_D = 70$ mÅ, and for a heliocentric angle $\theta \sim 84^\circ$ ($\cos \theta = 0.1$). The Einstein coefficient is $A = 6.27 \times 10^7$ s⁻¹, and the hyperfine structure constants for the Na atom are given in ref. 17. In **b**, the solid line shows the observations; the dashed line is the fit obtained with atomic polarization in the ground level (equation (3)). Values of the parameters are $[\beta_1]_0 = 2.0$, $[\beta_2]_0 = 3.8$, $\alpha = 0.1 k_c$, where k_c is the absorption coefficient of the Na atom at the centre of the D₂ line. See text for explanations of symbols.

collisions on the lower-level atomic polarization by supposing that this polarization is affected by depolarizing rates which increase exponentially with depth in the solar atmosphere. Assuming the same exponential behaviour for the absorption coefficient in the sodium lines, equation (2) becomes

$$Q/I = w \left(x_v + \frac{[x'_v]_0 - [x''_v]_0}{1 + \alpha/k_v} \right) \quad (3)$$

where $[x'_v]_0$ and $[x''_v]_0$ are the quantities x'_v and x''_v evaluated at the top of the solar atmosphere (where the alignment parameters are $[\beta_1]_0$ and $[\beta_2]_0$, respectively), k_v is the line absorption coefficient, and α is a constant related to the scale height of the solar atmosphere.

The polarization profile resulting from equation (3) is shown in Fig. 1b (dashed line). In spite of the approximations, the agreement between the theoretical result and the observed polarization profile is striking. Not only does the model explain the appearance of sharp polarization peaks in the cores of the D₁ and D₂ lines but it also reproduces quite well the three-lobed profile of D₂: I conclude that ground-level atomic polarization exists in the top layers of the solar atmosphere.

The existence of ground-level atomic polarization has important implications for solar physics. Because such polarization is much more sensitive to depolarizing mechanisms than upper-level atomic polarization, it can be used as a sensitive tool for diagnosing extremely weak magnetic fields through the ground-level Hanle effect. In my approximate model, ground-level atomic polarization turns out to be of the same order of magnitude (and indeed slightly larger) as the upper-level atomic polarization, implying the absence of depolarizing effects (both due to collisions and to magnetic fields) in the layers of the atmosphere where the cores of the D₁ and D₂ lines are formed. Although magnetic fields slightly inclined with respect to the vertical are not ruled out (a magnetic field of the order of 1 G and inclined by as much as 10–20° with respect to the vertical is compatible with the observations), the constraints on turbulent and/or canopy-like fields are much more severe. A turbulent field of the order of 1 G, for instance, would practically destroy ground-level atomic polarization, so that a polarization profile similar to the dashed line in Fig. 1a would result. The same is also true for horizontal fields of the same order with randomly distributed azimuth.

The results reported here seem to rule out the existence in the lower solar chromosphere of magnetic fields stronger than about 0.01 G either in the form of volume-filling, turbulent fields or in the form of canopy-like, horizontal fields. This contradicts previous evidence, supported both by observations^{3–8} and theory¹⁶, for existence of turbulent magnetic fields as strong as 10–20 G in the solar atmosphere outside active regions; however, this evidence mainly refers to the upper photosphere and to the layers where the 4,607-Å Sr I line is formed, ~200 km below the layers considered here. The results also contradict the evidence for the existence of canopy-like fields in the solar chromosphere^{9,10}.

A paradox therefore seems to exist; on the one hand the success of my theory in explaining the polarization profile of the sodium doublet demonstrates the presence of ground-level atomic polarization. On the other hand, this polarization cannot survive in the presence of turbulent or canopy-like fields stronger than about 0.01 G, which contradicts previous evidence. Resolution of this apparent paradox could be obtained if all the turbulent fields of the upper photosphere reconnect very efficiently at a level roughly corresponding to the temperature minimum, so that in the lower chromosphere only vertical (or quasi-vertical) fields survive. But this condition seems to be very *ad hoc* and implies the presence of some physical mechanism not yet understood. □

Received 12 September; accepted 29 December 1997.

1. Stenflo, J. O. & Keller, C. U. New window for spectroscopy. *Nature* **382**, 588 (1996).

2. Stenflo, J. O. & Keller, C. U. The second solar spectrum. A new window for diagnostics of the Sun. *Astron. Astrophys.* **321**, 927–934 (1997).
3. Stenflo, J. O. Small-scale magnetic structures on the Sun. *Astron. Astrophys. Rev.* **1**, 3–48 (1989).
4. Stenflo, J. O. *Solar Magnetic Fields* (Kluwer, Dordrecht, 1994).
5. Faurobert-Scholl, M. Investigation of microturbulent magnetic fields in the solar photosphere by their Hanle effect in the Sr I 4607 Å line. *Astron. Astrophys.* **268**, 765–774 (1993).
6. Faurobert-Scholl, M., Feautrier, N., Machefer, F., Petrovay, K. & Spielfiedel, A. Turbulent magnetic fields in the solar photosphere: diagnostics and interpretation. *Astron. Astrophys.* **298**, 289–302 (1995).
7. Stenflo, J. O., Keller, C. U. & Gandorfer, A. Differential Hanle effect and the spatial variation of turbulent magnetic fields on the Sun. *Astron. Astrophys.* **329**, 319–328 (1998).
8. Bianda, M., Solanki, S. K. & Stenflo, J. O. Hanle depolarisation in the solar chromosphere: The Ca I 4227 Å line. *Astron. Astrophys.* (in the press).
9. Jones, H. P. & Giovanelli, R. G. Magnetic canopies in unipolar regions. *Solar Phys.* **87**, 37–42 (1983).
10. Solanki, S. K. & Steiner, O. How magnetic is the solar atmosphere? *Astron. Astrophys.* **234**, 519–529 (1990).
11. Happer, W. Optical pumping. *Rev. Mod. Phys.* **44**, 169–249 (1972).
12. Landi Degl'Innocenti, E. The density matrix approach to polarized radiative transfer. *Solar Phys.* **164**, 21–28 (1996).
13. Trujillo Bueno, J. & Landi Degl'Innocenti, E. Linear polarization due to lower-level depopulation pumping in stellar atmospheres. *Astrophys. J.* **482**, L183–L186 (1997).
14. Landi Degl'Innocenti, E., Landi Degl'Innocenti, M. & Landolfi, M. in *Forum Themis* (eds Mein, N. & Sahal Bréchet, S.) 59–77 (Observatoire de Paris, 1996).
15. Shurcliff, W. A. *Polarized Light* (Harvard Univ. Press, Cambridge, Mass., 1966).
16. Petrovay, K. & Szakály, G. The origin of intranetwork fields: a small-scale solar dynamo. *Astron. Astrophys.* **274**, 543–554 (1993).
17. Landolfi, M. & Landi Degl'Innocenti, E. Polarization of the sodium D lines in prominences. *Solar Phys.* **98**, 53–66 (1985).

Acknowledgements. I thank J. O. Stenflo for helping in the presentation of the Letter and for providing the original data shown in Fig. 1.

Correspondence and requests for materials should be addressed to E.L.D. (e-mail: landie@arcetri.astro.it).

Tunnelling and zero-point motion in high-pressure ice

Magali Benoit*, Dominik Marx† & Michele Parrinello†

* Laboratoire des Verres, Université Montpellier II, 34095 Montpellier, France

† Max-Planck-Institut für Festkörperforschung, Heisenbergstrasse 1, 70569 Stuttgart, Germany

The microscopic structure of ice poses a long-standing challenge to theory^{1–3}. Because of their low mass, the protons in the hydrogen bonds that define the structures of crystalline ice are susceptible to quantum-mechanical effects such as tunnelling^{1,4–8}. High pressure provides a means of controlling the length of the hydrogen bonds in order to investigate such effects. In particular, Holzapfel predicted 26 years ago that, under pressure, hydrogen bonds might be transformed from the highly asymmetric O–H···O configuration to a symmetric state in which the proton lies midway between the two oxygens⁹, leading to a non-molecular symmetric phase of ice, now denoted as ice 'X'. The existence of this phase has been inferred from spectroscopy^{10–14}, but has still not been observed directly. Here we investigate the role of quantum effects in proton ordering and hydrogen-bond symmetrization within ice at high pressure by using a simulation technique that treats both electrons and nuclei quantum-mechanically^{15–17}. We find that the proton-ordered structure at low pressure, with asymmetric hydrogen bonds (ice VIII), transforms on increasing pressure to a proton-disordered asymmetric phase (ice VII) owing to translational proton tunnelling. On further compression, the zero-point fluctuations lead to strongly delocalized protons and hydrogen-bond symmetrization, even though the underlying character of the proton-transfer potential remains a double well. Only at still higher pressures does the double-well potential become transformed into a single well, whereupon the protons again become increasingly localized.

The rich phase diagram of ice reduces to two known molecular ice phases above about 2 GPa, ice VIII and VII (ref. 18). Ice VII is best understood as being composed of two interpenetrating but not interconnected tetrahedral hydrogen-bonded networks of water molecules of normal cubic ice, I_c, with a body-centred-cubic

(b.c.c.) oxygen structure. The orientation of the water molecules is random, apart from short-range order imposed by obeying the local 'ice rules'¹⁸. Contrary to this paraelectric phase VII, the water molecules and the associated dipole moments on the two sublattices do possess long-range order and point in opposite directions in the antiferroelectric phase VIII; the dipolar ordering is accompanied by a slight tetragonal distortion of the cubic unit cell along the staggered polarization axis¹⁹. The antiferroelectric disordering transition is characterized by a significant and constant H/D isotope effect of about 10 GPa below ~ 100 K, where the transition pressure (62 GPa for H₂O and 72 GPa for D₂O) is independent of temperature^{10,11}.

In the conjectured symmetric ice X (ref. 9), the oxygen lattice remains b.c.c., while the protons reside midway between their two nearest oxygen neighbours, instead of forming one covalent O–H and one H $\cdots$ O bond as in all low-pressure molecular phases such as ice VIII and VII. Despite numerous claims in the literature to have detected ice X, so far only infrared experiments have provided evidence for a transition towards a symmetrized structure at 60–70 GPa (~ 10 GPa higher for D₂O) in the range from room temperature down to at least 80 K (refs 12–14). However, the spectra are found to be congested as a result of various resonance phenomena^{14,20,21}, rendering their interpretation difficult. The symmetrization issue has been given a new twist by the idea^{22–24} that two such forms might exist, namely 'proton-disordered symmetric ice' and 'proton-ordered symmetric ice'; see also refs 12, 14.

Here we approach these open questions based on our *ab initio* path integral technique^{15–17}, which satisfies all essential requirements posed by this complex system. Briefly, we treat all nuclei fully quantum-mechanically using path integrals²⁵ without reducing the dimensionality⁴ or invoking other approximations⁴ such as the

quasiclassical or quasiharmonic ones^{7,8}, except that of neglecting the exchange of identical particles. The forces on the nuclei are derived for every individual configuration within the Born–Oppenheimer approximation from first-principles electronic structure calculations in the framework of Hohenberg–Kohn–Sham density functional theory²⁶, allowing chemical bonds to be broken and made. In view of the computational costs involved, no correction for the finiteness of the system, such as finite-size scaling and Trotter extrapolation techniques²⁵, could be applied. The conversion of the simulation cell volume to pressure is based on the standard experimental equation of state²⁷. Thus, the pressure scale including the reported approximate transition pressures cannot be considered as precise numbers.

The degree of the antiferroelectric order is measured as the difference between the two sublattice polarizations of ice VIII. Our order parameter D (see Fig. 1 legend for definition) is about 1.2 qÅ in ideal ice VIII, whereas it drops to zero in an infinitely large 'perfect' paraelectric ordered structure. The fully quantum-mechanical simulations at 100 K (Fig. 1, filled circles) clearly start out in ice VIII at our lowest pressure. On compression, the order parameter stays close to the ideal value 1.2 qÅ but drops significantly at ~ 47 GPa down to a plateau value of $\sim 0.4 \text{ qÅ}$ up to the highest pressure. The simulation with classical nuclei at 100 K (Fig. 1, open circles) follows a similar pattern, but the antiferroelectric disordering now takes place at nearly twice the pressure (~ 76 GPa). A classical reference calculation at 300 K in the disordered phase VII (Fig. 1, squares) shows that the residual order of $\sim 0.4 \text{ qÅ}$ is due to local short-range correlations imposed by the ice rules¹⁸ in conjunction with the limited system size and the periodic boundary conditions.

The symmetrization of the hydrogen bonds is best followed in

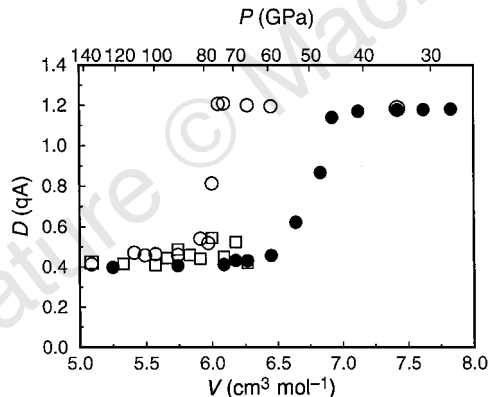


Figure 1 Antiferroelectric staggered order parameter D as a function of the molar volume V . We define D as the difference between the two sublattice polarizations 'up' and 'down', $D = \langle (1/N) | \sum_{j \in \text{up}} \mathbf{d}_j - \sum_{j \in \text{down}} \mathbf{d}_j | \rangle$, where $\mathbf{d}_j$ is the purely geometric dipole moment vector of the j th water molecule obtained with the formal charges $-2q$ and $+1q$ on oxygen and hydrogen atoms, respectively. Filled circles: quantum nuclei at 100 K; open circles: classical nuclei at 100 K; open squares: classical nuclei at 300 K. The conversion of the simulation cell volume to pressure is based on the experimental H₂O equation of state at 300 K (ref. 27); note that the H/D isotope effect on this quantity is as small as the experimental uncertainty²⁴. All calculations were made with 16 H₂O molecules in a periodic cubic cell relying on a gradient-corrected³¹ local density approximation³², Troullier–Martins pseudopotentials³³ and a Γ -point plane wave expansion of the valence orbitals up to 70 Rydberg. The path integrals²⁵ for the quantum simulations^{15,16} were discretized in imaginary time using 16 replica. The staging propagator together with one chain of three Nosé–Hoover thermostats per degree of freedom was used to improve sampling efficiency, ensure ergodicity and generate the canonical ensemble¹⁷. After careful equilibration, statistical averages were sampled from about 10,000 and 30,000 configurations in the quantum and classical cases, respectively.

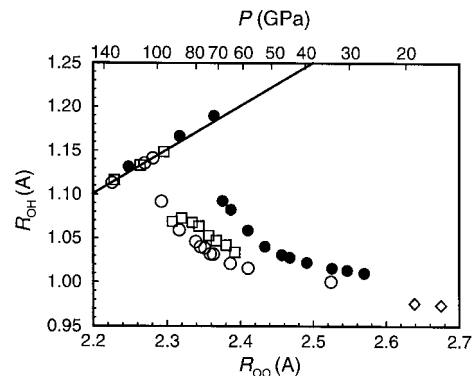


Figure 2 Evolution of the average oxygen-hydrogen bond length R_{OH} as a function of the average oxygen-oxygen distance R_{OO} . Filled circles: quantum nuclei at 100 K; open circles: classical nuclei at 100 K; open squares: classical nuclei at 300 K; open diamonds: experimental (R_{OO} , R_{OH}) data at about 240 K (ref. 28); solid line: 'symmetrization line' $R_{\text{OH}} = R_{\text{OO}}/2$, indicating ice X. For pressure conversion, see Fig. 1; note that as we plot the experimental data directly as (R_{OO} , R_{OH}) pairs, their corresponding pressure in this figure results from our pressure conversion according to ref. 27.

Fig. 2 by monitoring the evolution of the quantum-statistical average of the shortest oxygen–hydrogen bond length R_{OH} as a function of the nearest-neighbour oxygen–oxygen distance R_{OO} (which in turn is a function of pressure), the signature of symmetric ice simply being $R_{OH} = R_{OO}/2$ (solid line). We find this relation to be approximately linear in the low pressure regime, which has also been found by neutron scattering up to 10 GPa (ref. 19) and 20 GPa (ref. 28, and S. Klotz, personal communication). On compression, however, we find that the lengthening of the R_{OH} bond becomes much more pronounced as a function of decreasing R_{OO} . The quantum system (Fig. 2, filled circles) finally undergoes a symmetrization transition at ~ 72 GPa, whereas in the corresponding classical system (Fig. 2, open circles) a much higher pressure of ~ 102 GPa is needed to break the molecules apart. So far we have found that at ~ 100 K, phase VIII does not transform immediately into symmetric ice: there is clear evidence that proton-disordered molecular ice VII exists between proton-ordered ice VIII at lower pressures and symmetric non-molecular ice X at higher pressures.

At this stage we can go beyond experiment by unravelling the physical mechanisms that lead to the observed isotope effects by comparing the quantum simulations to the classical ones under the

same conditions; furthermore, such a comparison yields theoretical upper bounds for the H/D isotope effects on the phase transitions, as quantified in Figs 1 and 2. To this end, we have analysed (Fig. 3) the average proton distribution $P(\delta, R_{O_aO_b})$ as a function of the proton position relative to the bond midpoint $\delta = R_{O_aH} - R_{O_bH}$ and the corresponding oxygen–oxygen separation $R_{O_aO_b}$. The quantum simulations shown in Fig. 3a and b correspond to ice VIII and VII, respectively, whereas those in Fig. 3c, d correspond to ice X. In ice VIII (Fig. 3a) the quantum fluctuations only induce a zero-point motional broadening of the distribution relative to the classical case (Fig. 3e), without affecting its shape. The quantum effects in ice VII (Fig. 3b) are dramatic, where it is proton tunnelling along the hydrogen bonds, absent in the classical case Fig. 3f, that leads to proton disorder and thus to a bimodal distribution. Finally, in ice X (Fig. 3c) we find a broad and flat but unimodal distribution centred at $\delta = 0$, which sharpens upon further compression (Fig. 3d). In contrast, the classical calculation (Fig. 3g) is characterized by a bimodal distribution, reflecting an underlying double-well proton-transfer effective potential. Taking into account that the thermal fluctuations correspond to the same temperature in Fig. 3c and g, it is a zero-point motion effect which shifts the lowest vibrational level of the proton wavefunction above the double-well barrier. This form of ice (Fig. 3c) could be dubbed ‘proton-disordered symmetric ice’ with delocalized protons because the proton distribution function is very broad, covers both potential wells, but peaks at the bond midpoint, $\delta = 0$. Only at still higher compression of ≥ 102 GPa (from Fig. 2) do we observe a single well (Fig. 3h) with the potential minimum right at the centred position $\delta = 0$, which in turn could be called ‘proton-ordered symmetric ice’ (Fig. 3d) with localized protons at the bond midpoints; see refs 12, 14, 22–24 for earlier conjectures along these lines. Thus, we trace back the measured H/D isotope effects on the ice VIII $\rightarrow$ VII and symmetrization transitions to proton tunnelling along the hydrogen bonds and zero-point motion, respectively.

From our path integral simulations, we can estimate the spread of the proton wavefunction (see Fig. 4 legend for a precise definition)

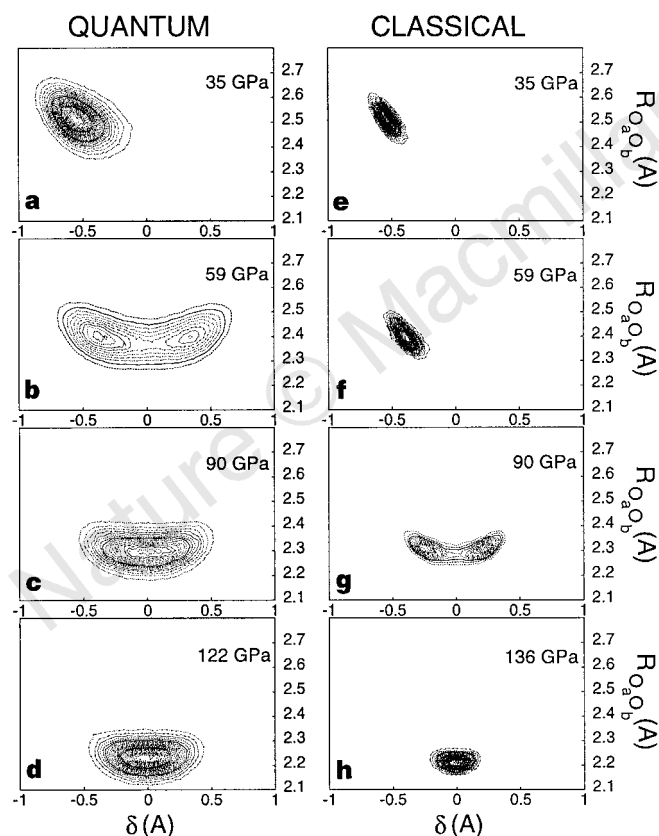


Figure 3 Proton symmetrization as a function of shortening of the hydrogen bond. Here we show contour plots of the average proton distribution function $P(\delta, R_{O_aO_b})$ as a function of the proton position relative to the bond midpoint $\delta = R_{O_aH} - R_{O_bH}$ and the corresponding oxygen–oxygen separation $R_{O_aO_b}$ for quantum (left panels) and classical (right panels) simulations at several representative volumes and 100 K; note that $\delta = 0$ corresponds to the proton being located midway between its neighbouring oxygen atoms O_a and O_b , irrespective of the actual $R_{O_aO_b}$ distance. Quantum: **a**, $V = 7.42 \text{ cm}^3 \text{ mol}^{-1}$ (~ 35 GPa); **b**, $V = 6.45 \text{ cm}^3 \text{ mol}^{-1}$ (~ 59 GPa); **c**, $V = 5.74 \text{ cm}^3 \text{ mol}^{-1}$ (~ 90 GPa); **d**, $V = 5.25 \text{ cm}^3 \text{ mol}^{-1}$ (~ 122 GPa); classical: **e**, $V = 7.42 \text{ cm}^3 \text{ mol}^{-1}$ (~ 35 GPa); **f**, $V = 6.45 \text{ cm}^3 \text{ mol}^{-1}$ (~ 59 GPa); **g**, $V = 5.74 \text{ cm}^3 \text{ mol}^{-1}$ (~ 90 GPa); **h**, $V = 5.08 \text{ cm}^3 \text{ mol}^{-1}$ (~ 136 GPa). For pressure conversion, see Fig. 1.

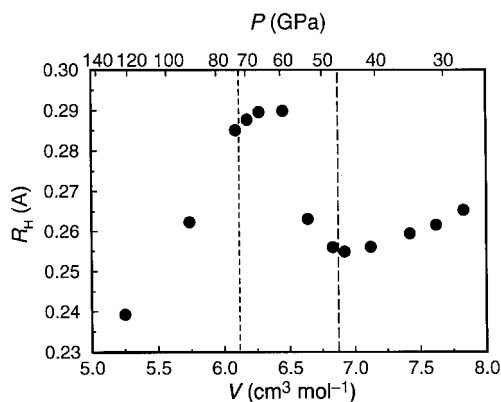


Figure 4 Evolution of the average spread of the quantum protons R_H as a function of the molar volume V at 100 K. We define the correlation function²⁹ $R_H(\Delta\tau) = \Sigma_{i=1}^N \langle |\mathbf{R}_i(\tau) - \mathbf{R}_i(\tau')|^2 \rangle^{1/2} / N$, where $0 \leq \Delta\tau = \tau - \tau' \leq \hbar\beta$, $\beta = 1/k_B T$, $\mathbf{R}_i(\tau)$ denotes the position of the i th proton at imaginary time τ ; note that $R_H(\Delta\tau) = 0 \text{ Å}$ for classical protons. The midpoint value $R_H = R_H(\hbar\beta/2)$ is a measure of the particles' spatial extents, whereas its variation in imaginary time (not shown here) determines the degree of ground-state dominance and thus localization²⁹. Localized protons with a large energy gap ΔE between ground and excited states are characterized by a constant function in the range of roughly $\hbar\Delta E < \Delta\tau < \hbar\beta - \hbar\Delta E$, whereas $R_H(\Delta\tau)$ varies in the entire $\Delta\tau$ interval for extended or delocalized protons. The long-dash and short-dash vertical lines mark the VIII $\rightarrow$ VII and VII $\rightarrow$ X transitions, respectively, as a function of increasing pressure. For pressure conversion, see Fig. 1.

and get a glimpse of the structure of the proton excited states^{29,30}. The results depicted in Fig. 4 as a function of pressure further strengthen our conclusions. Upon compressing ice VIII, we find that the protons become increasingly compact, ground-state dominated and thus localized. Overstepping the phase boundary to ice VII (long dashed line), their spatial extent increases rapidly and is accompanied by a strong delocalization: that is, extended protonic states contribute significantly in addition to the ground state. Such behaviour is also characteristic immediately after crossing the phase transition to ice X (short dashed line), but there is a clear trend for protons to become strongly localized again at the highest pressure, where the underlying potential is of the single-well type.

Our essential findings can be summarized as follows: (1) The sequence of the transitions is antiferroelectric ice VIII to paraelectric ice VII and finally to symmetric ice X at about 100 K; (2) the isotope effect on the antiferroelectric disordering transition (VIII $\rightarrow$ VII) is caused by translational proton tunnelling, whereas that on symmetrization (VII $\rightarrow$ X) is a qualitative zero-point motion effect; and (3) there are two 'forms' of ice X, both characterized by a unimodal proton distribution centred at the bond midpoint, the one at lower pressure being characterized by an underlying double-well proton transfer potential and a very broad proton distribution, whereas the high-pressure form has a single-well potential with a sharper proton localization. We hope that the insight we have gained from the different regimes of a 'clean' and relatively simple system, ice under controlled high pressure, will be useful in understanding quantum effects in other proton-transfer systems. □

Received 15 October 1997; accepted 7 January 1998.

- Bernal, J. D. & Fowler, R. H. A theory of water and ionic solutions, with particular reference to hydrogen and hydroxyl ions. *J. Chem. Phys.* **1**, 515–548 (1933).
- Pauling, L. The structure and entropy of ice and of other crystals with some randomness of atomic arrangement. *J. Am. Chem. Soc.* **57**, 2680–2684 (1935).
- Bjerrum, N. Structure and properties of ice. *Science* **115**, 385–390 (1952).
- Schweizer, K. S. & Stillinger, F. H. High pressure phase transitions and hydrogen-bond symmetry in ice polymorphs. *J. Chem. Phys.* **80**, 1230–1240 (1984).
- McMahon, M. I. *et al.* Geometric effects of deuteration on hydrogen-ordering phase transitions. *Nature* **348**, 317–319 (1990).
- Krumhansl, J. A. Sorting chickens from eggs. *Nature* **348**, 285–286 (1990).
- Lee, C., Vanderbilt, D., Laasonen, K., Car, R. & Parrinello, M. *Ab initio* studies on high pressure phases of ice. *Phys. Rev. Lett.* **69**, 462–465 (1992).
- Lee, C., Vanderbilt, D., Laasonen, K., Car, R. & Parrinello, M. *Ab initio* studies on the structural and dynamical properties of ice. *Phys. Rev. B* **47**, 4863–4872 (1993).
- Holzappel, W. B. On the symmetry of the hydrogen bonds in ice VII. *J. Chem. Phys.* **56**, 712–715 (1972).
- Pruzan, P., Chervin, J. C. & Canny, B. Stability domain of the ice VIII proton-ordered phase at very high pressure and low temperature. *J. Chem. Phys.* **99**, 9842–9846 (1993).
- Pruzan, P. Pressure effects on the hydrogen bond in ice up to 80 GPa. *J. Mol. Struct.* **322**, 279–286 (1994).
- Goncharov, A. F., Struzhkin, V. V., Somayazulu, M. S., Hemley, R. J. & Mao, H. K. Compression of ice to 210 GPa: Infrared evidence for a symmetric hydrogen-bonded phase. *Science* **273**, 218–220 (1996).
- Aoki, K., Yamawaki, H., Sakashita, M. & Fujihisa, H. Infrared absorption study of the hydrogen-bond symmetrization in ice to 100 GPa. *Phys. Rev. B* **54**, 15673–15677 (1996).
- Struzhkin, V. V., Goncharov, A. F., Hemley, R. J. & Mao, H. K. Cascading Fermi resonances and the soft mode in dense ice. *Phys. Rev. Lett.* **78**, 4446–4449 (1997).
- Marx, D. & Parrinello, M. *Ab initio* path-integral molecular dynamics. *Z. Phys. B (Rapid Note)* **95**, 143–144 (1994).
- Marx, D. & Parrinello, M. *Ab initio* path integral molecular dynamics: Basic ideas. *J. Chem. Phys.* **104**, 4077–4082 (1996).
- Tuckerman, M. E., Marx, D., Klein, M. L. & Parrinello, M. Efficient and general algorithms for path integral Car–Parrinello molecular dynamics. *J. Chem. Phys.* **104**, 5579–5588 (1996).
- Hobbs, P. V. *Ice Physics* (Clarendon Press, Oxford, 1974).
- Nelmes, R. J. *et al.* Neutron diffraction study of the structure of deuterated ice VIII to 10 GPa. *Phys. Rev. Lett.* **71**, 1192–1195 (1993).
- Aoki, K., Yamawaki, H. & Sakashita, M. Pressure-tuned Fermi Resonance in ice VII. *Science* **268**, 1322–1324 (1995).
- Aoki, K., Yamawaki, H. & Sakashita, M. Observation of Fano interference in high-pressure ice VII. *Phys. Rev. Lett.* **76**, 784–786 (1996).
- Hama, J. & Suito, K. Evidence of a new phase of ice above 70 GPa from the analysis of experimental data using the universal equation of state. *Phys. Lett. A* **187**, 346–350 (1994).
- Pruzan, P. *et al.* Raman scattering and X-ray diffraction of ice in the Megabar range. Occurrence of a symmetric disordered solid above 62 GPa. *J. Phys. Chem. B* **101**, 6230–6233 (1997).
- Wolaniec, E. *et al.* Equation of state of ice VII up to 106 GPa. *Phys. Rev. B* **56**, 5781–5785 (1997).
- Ceperley, D. M. Path integrals in the theory of condensed helium. *Rev. Mod. Phys.* **67**, 279–355 (1995).
- Jones, R. O. & Gunnarsson, O. The density functional formalism, its applications and prospects. *Rev. Mod. Phys.* **61**, 689–746 (1989).
- Hemley, R. J. *et al.* Static compression of H₂O-ice to 128 GPa (1.28 Mbar). *Nature* **330**, 737–740 (1987).
- Nelmes, R. J., Loveday, J. S., Marshall, W. G., Besson, J. M., Klotz, S. & Hamel, G. Structures of ice VII and ice VIII to 20 GPa. *Proceedings of the International Conference on High Pressure Science and Technology (Joint Conference: AIRAPT-16 & HPCJ-38)* (Kyoto, August 25–29, 1997).

- Chandler, D. & Leung, K. Excess electrons in liquids, geometrical perspectives. *Annu. Rev. Phys. Chem.* **45**, 557–591 (1994).
- Stich, I., Marx, D., Parrinello, M. & Terakura, K. Proton-induced plasticity of hydrogen clusters. *Phys. Rev. Lett.* **78**, 3669–3672 (1997).
- Becke, A. D. Density-functional exchange–energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).
- Perdew, J. P. & Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23**, 5048–5079 (1981).
- Troullier, N. & Martins, J. L. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B* **43**, 1993–2006 (1991).

Acknowledgements. Our warm thanks to M. Tuckerman, J. Hutter, U. Schwarz, M. Bernasconi and S. Klotz for useful discussions. The simulations were carried out on the IBM SP2 at CNUSC (Montpellier) and at MPI Stuttgart.

Correspondence should be addressed to D.M. (marx@prr.mpi-stuttgart.mpg.de).

Polarizing energy transfer in photoluminescent materials for display applications

Andrea Montali, Cees Bastiaansen, Paul Smith & Christoph Weder

Department of Materials, Institute of Polymers, ETH Zürich, CH-8092 Zürich, Switzerland

Combinations of sheet polarizers and colour filters form the basis of numerous products^{1–4}—most notably colour liquid-crystal displays^{2,4}—that require polarized chromatic light. But this combination of elements does not make efficient use of light, as a substantial fraction of the incident light is converted into thermal energy^{3,4}, limiting the brightness and energy efficiency of the resulting devices. Here we show that these limitations can be overcome by using polymer-based photoluminescent polarizers. Our polarizers operate on a two-step principle: randomly orientated 'sensitizer' molecules harvest the incident light by isotropic absorption and then efficiently transfer the energy to a uniaxially orientated photoluminescent polymer, from which coloured light with a high degree of linear polarization is emitted. In principle, isotropic-to-polarized conversion efficiencies approaching unity could be attainable by this approach.

The present limitations of liquid-crystal displays (LCDs) have recently triggered the development of polarizers based on selective reflection or scattering of light^{5–9} which can replace the dichroic polarizer in a conventional LCD configuration. Using appropriate supplementary elements to recycle reflected or scattered energy, the ultimate efficiency of these systems can, in principle, approach unity (thus, twice that of dichroic polarizers); but in practice the efficiency is up to 80% (refs 5–9). However, colour applications require filters, which absorb at least two-thirds of the light⁴, and thus reduce the ultimate overall efficiency to below 30%. In another recent approach, photoluminescent (PL) polarizers were shown to combine efficiently the polarization of light and the production of bright colours^{10,11}, together with a substantial increase in brightness and efficiency of PL LCDs based on these elements^{12,13}. These polarizers comprise uniaxially orientated PL polymers, which, after photoexcitation, emit linearly polarized light. Their efficiency is chiefly limited by the luminophore's quantum yield which, ultimately, can approach unity, but for PL polymers is typically up to 80% (refs 14, 15). However, when used in a standard PL LCD configuration¹², only ~50% of light incident from the light source is used, as the absorption of these PL polarizers is also anisotropic^{10,11}.

We report here materials that show nearly isotropic absorption, but which emit the absorbed energy in highly polarized fashion, and, consequently, allow the production of PL polarizers that approach the ultimate efficiency. Such polarizers can directly replace the standard polarizer in conventional LCDs and, using an appropriate (ultraviolet) backlight and a dichroic mirror (to direct

all emitted light towards the viewer), result in ultra-efficient, coloured PL LCDs. The thermoplastic character of these materials, in principle, also allows the production of patterned structures, which enable multicolour devices. Although the reverse effect (that is, PL depolarization) is well known^{16–18}, the polarizing energy transfer shown by the new materials is not only of technological relevance but also manifests a new photophysical phenomenon.

The PL films we report here are based on uniaxially orientated,

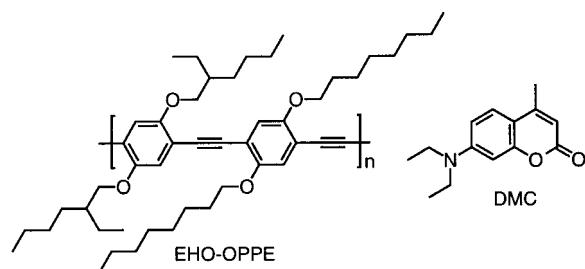


Figure 1 Molecular structures of compounds used. Left, the poly(2,5-dialkoxy-*p*-phenylene ethynylene) derivative (EHO-OPPE, number-average molecular weight $\sim 1 \times 10^4 \text{ g mol}^{-1}$; right, 7-diethylamino-4-methylcoumarin (DMC; Aldrich).

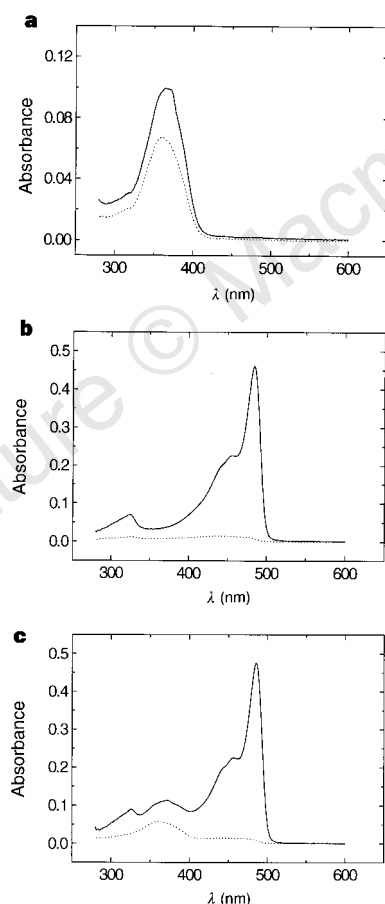


Figure 2 Polarized absorption spectra of orientated films. Spectra were obtained with *p*- (solid line) and *s*- (dashed line) polarized light (see text for definition of *s* and *p*). **a**, Binary UHMW PE/DMC blend; **b**, Binary UHMW PE/EHO-OPPE blend; **c**, ternary UHMW PE/EHO-OPPE/DMC blend. UHMW PE (weight-average molecular weight $\sim 4 \times 10^6 \text{ g mol}^{-1}$) was from Hoechst AG. Spectra were recorded on a Perkin Elmer $\lambda 900$; films were sandwiched with silicone oil between quartz slides for all experiments to minimize light-scattering at film surfaces.

ternary blends of ultra-high-molecular-weight polyethylene (UHMW PE), a poly(2,5-dialkoxy-*p*-phenylene ethynylene) derivative (EHO-OPPE)¹⁵ (2% w/w), and 7-diethylamino-4-methylcoumarin (DMC) (2% w/w) as the sensitizer (Fig. 1). The respective binary blends (UHMW PE/EHO-OPPE and UHMW PE/DMC) were used as reference systems. Blend films were prepared by solution-casting of EHO-OPPE (10 mg), and/or DMC (10 mg), and UHMW PE (500 mg) from xylene (50 g) according to standard procedures¹¹. The resulting films were uniaxially drawn at 120 °C to draw ratios of ~ 80 , yielding orientated PL films of a thickness of 1–2 μm and of low optical density.

DMC was selected as the sensitizer because of its low form-anisotropy, suitable photophysical prerequisites and particularly beneficial phase behaviour. The melting temperature of 74 °C makes DMC compatible with the orientation process which requires mobility of the guest molecules during deformation; in addition, DMC and EHO-OPPE are miscible at elevated temperatures, which enables a most favourable morphology of the orientated blends (see below). The absorption of DMC at a wavelength of $\sim 364 \text{ nm}$ optimally overlaps with the emission of common ultraviolet lamps that may be used as excitation sources in PL LCDs^{12,13}. Importantly, DMC seems not to quench emission of EHO-OPPE and, mandatory for energy transfer¹⁹, its own emission favourably overlaps with the absorption of EHO-OPPE.

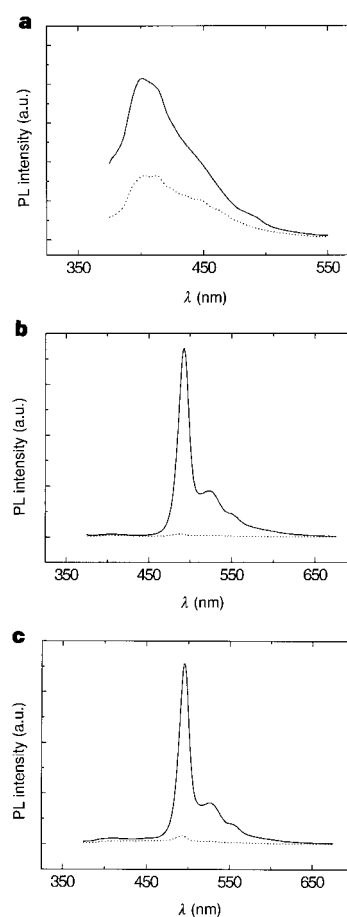


Figure 3 Emission spectra of orientated films obtained under isotropic excitation at 365 nm. Polarized detection in *p*- (solid line) and *s*- (dashed line) mode was used. **a**, Binary UHMW PE/DMC blend; **b**, binary UHMW PE/EHO-OPPE blend; **c**, ternary UHMW PE/EHO-OPPE/DMC blend. Emission spectra were recorded on a SPEX Fluorolog F212; corrected PL intensities are given in arbitrary units, but spectra compared in one graph always have identical scale.

We investigated photophysical characteristics of PL films based on the ternary and the binary reference blends, using polarized ultraviolet-visible absorption and steady-state PL spectroscopy. As the absorption at 440 nm is exclusively related to EHO-OPPE, and the absorption at 365 nm is principally due to DMC, experiments

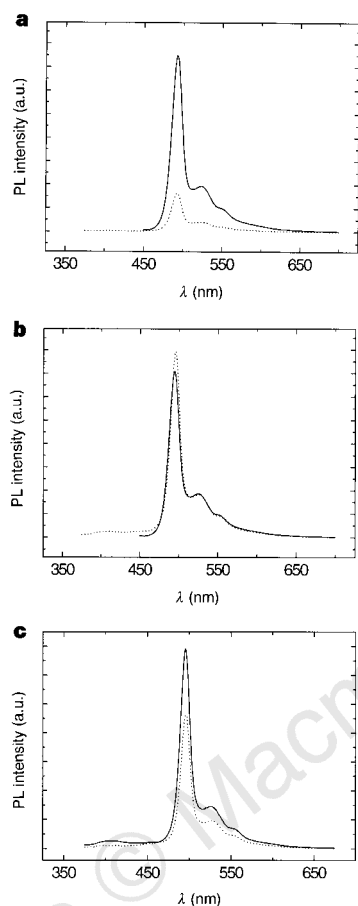
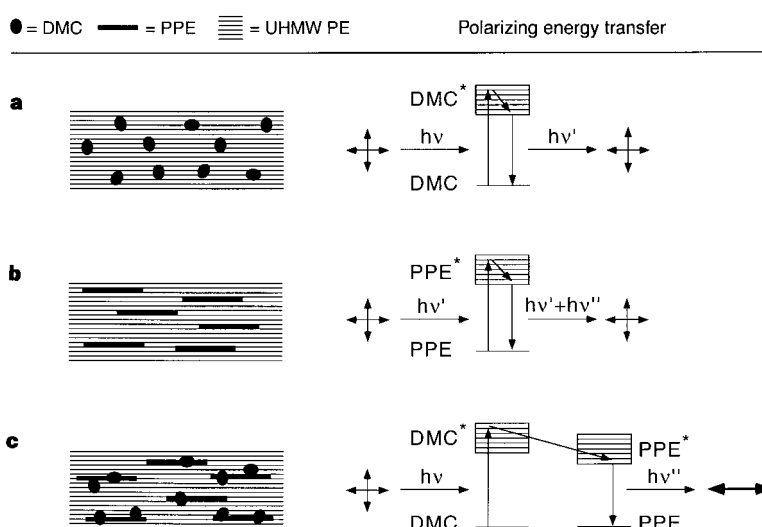


Figure 4 Emission spectra of orientated films. **a** and **b**, emission spectra of orientated films obtained under isotropic excitation at 440 (solid line) and 365 nm (dashed line), and polarized detection in p-mode; **a**, binary blend of UHMW PE/EHO-OPPE; **b**, ternary blend of UHMW PE/EHO-OPPE/DMC. **c**, Emission spectra of an orientated UHMW PE/EHO-OPPE/DMC ternary blend film obtained under polarized excitation at 365 nm in p- (solid line) and s- (dashed line) mode and polarized detection in p-mode.

Figure 5 Photophysical processes and optical emission. **a–c**, Schematic representation of the photophysical processes observed for uniaxially orientated films of the binary reference blends of UHMW PE/DMC (**a**), UHMW PE/EHO-OPPE (**b**) and the ternary UHMW PE/EHO-OPPE/DMC blend (**c**). Arrows indicate polarizations of incident and emitted light.



were performed at these two wavelengths to separately address the conjugated polymer and the sensitizer. (We note that different phonon bands are observed for EHO-OPPE for p- and s-polarized light (see below), making the dichroic behaviour wavelength-dependent, with maximum distortion at absorption maxima; to better reflect the 'average' dichroic behavior, expressed by the division of the integrated spectra¹¹, experiments to address EHO-OPPE were conducted at the empirically determined wavelength of 440 nm.)

Polarized absorption spectra acquired with incident light polarized parallel (p) and perpendicular (s) to the orientation direction of the films, show (Fig. 2) that the absorption characteristics of the ternary blend are a combination of those of the two respective binary blends. The ternary blend has high absorption dichroic ratios DR_A (ratio between absorption for p- and s-polarized light) of up to 13 at 440 nm, resulting from a high degree of orientation of EHO-OPPE. By contrast, the absorption at 365 nm is essentially isotropic ($DR_A = 1.5$) and reflects the nearly random orientation of the sensitizer within the orientated UHMW PE matrix.

Polarized emission spectra, obtained under isotropic excitation at 365 nm and polarized detection in either p- or s-mode, are shown in Fig. 3. In binary UHMW PE/DMC films, the emission from DMC, centred around 400 nm, shows only minor polarization, expressed by an emission dichroic ratio, DR_E (ratio between the integrated PL spectra in p- and s-mode), of 2.3, consistent with the low degree of orientation of the sensitizer. By contrast, the binary UHMW PE/EHO-OPPE films show state-of-the-art emission anisotropy¹¹ ($DR_E = 27$). In the ternary blend, importantly, the DMC emission is almost fully suppressed, while the emission from EHO-OPPE is highly polarized ($DR_E = 16$). The fact that DR_E is somewhat lower in the ternary than in the binary UHMW PE/EHO-OPPE blend is explained by the plasticizing effect of DMC on EHO-OPPE which somewhat reduces the efficiency of the orientation process.

Energy transfer from DMC to the conjugated polymer is evident when comparing the emission intensities (related to EHO-OPPE) of the ternary and the binary UHMW PE/EHO-OPPE blend (Fig. 4a, b) for isotropic excitation at 440 and 365 nm, respectively. The binary reference blend shows a significantly lower emission intensity when excited at 365 nm compared to excitation at 440 nm, due to the much lower absorption of EHO-OPPE at the shorter wavelength (Fig. 2b). The ternary blend, by contrast, shows similar emission intensities when excited at 365 and 440 nm, as a result of the sensitizing effect of DMC: the effective, isotropic absorption of the sensitizer, evidently followed by energy transfer to the conjugated polymer, is the obvious rationalization for the increased emission intensity. The polarizing characteristic of the energy transfer is shown by the results given in Fig. 4c. The intensity of

p-polarized emission from the ternary blend was found to be only weakly dependent on the polarization of the incident light (when excited at 365 nm). In fact, the ratio of the emission intensities for excitation with s- and p-polarized light (1.5) is in gratifying agreement with the slightly dichroic absorption of the film at 365 nm ($DR_A = 1.5$). Thus, the ternary blend unambiguously shows the phenomenon of polarizing energy transfer: optical energy is nearly isotropically absorbed by DMC, and, with similar efficiency for both absorption (excitation) polarizations, transferred to EHO-OPPE, which subsequently emits polarized light. In the most unfavourable limit (Fig. 4c) the new material converts fully s-polarized into highly p-polarized light.

The polarizing energy transfer process observed in the present PL materials is schematically represented in Fig. 5a–c, together with the respective processes in the binary reference systems. Principally, the phenomenon may originate from either radiative¹⁹ (trivial), long-range coulombic²⁰ (Förster) or short-range electron-exchange²¹ (Dexter) energy transfer between the DMC sensitizer as donor and the orientated EHO-OPPE as acceptor. The fact that energy is transferred between donor molecules that have been excited with s-polarized light and acceptor molecules which subsequently emit p-polarized light implies a depolarization of the donor excited state, unless Dexter-type coupling is involved¹⁹. This depolarization could be derived from randomizing energy migration²² or orientational relaxation of the donor¹⁶ and is indeed observed when exciting the binary DMC reference blend with polarized light. The low optical density of the samples essentially excludes radiative energy transfer^{16–18}. A non-radiative energy transfer might, on the other hand, point to a very particular phase behaviour of the orientated blends. We have shown earlier (for binary blends) that EHO-OPPE forms an apparent molecular dispersion in the UHMW PE matrix¹¹. Thus, the incompatibility of DMC and UHMW PE, and the demonstrated affinity of DMC and EHO-OPPE make the formation of DMC/EHO-OPPE aggregates very likely, in which a non-radiative energy transfer is enabled by the close proximity of donor and acceptor molecules. Further experiments, including time-resolved measurements and the determination of quantum efficiencies, are in progress to develop of a full understanding of the underlying mechanism.

As a direct indication of the practical usefulness of DMC sensitization, we measured absolute brightness under isotropic excitation with a 365-nm ultraviolet lamp—a configuration of relevance to actual PL LCD^{12,13}. The luminosity of a ternary blend film is dramatically increased (82 cd m⁻²), compared to an unsensitized binary blend (22 cd m⁻²) of similar optical density in the EHO-regime. Of course, the absolute brightness could be further enhanced by an increase in optical density.

To explore other compositions of this new class of multifunctional materials, we prepared films containing an alternative sensitizer (7-[dimethylamino]-2,3-dihydrocyclopenta[c][1]benzopyran-4[1H]-one) and an alternative acceptor (poly[2-methoxy-5-[2'-ethyl-hexyloxy]-p-phenylenevinylene], MEH-PPV; ref. 23). These systems showed energy-transfer characteristics similar to those of the present ternary blend and, thus, demonstrate the versatility of the concepts outlined here²⁴. □

Received 11 November 1997; accepted 19 January 1998.

1. Klier, D. S., Lewis, J. W. & Randall, C. E. *Polarized Light in Optics and Spectroscopy* (Academic, San Diego, 1990).
2. Chan, L. K. M. in *The Encyclopedia of Advanced Materials* Vol. 2 (eds Bloor, D., Brook, R. J., Flemings, M. C. & Mahajan, S.) 1294–1304 (Elsevier Science, Oxford, 1994).
3. Draica, P. J. *Liquid Crystal Dispersions* (World Scientific, Singapore, 1995).
4. Nelson, T. J. & Wullert, J. R. III *Electronic Information Display Technologies* (World Scientific, Singapore, 1997).
5. Schadt, M. & Fünfschilling, J. New liquid crystal polarized color projection principle. *Jpn J. Appl. Phys.* **29**, 1974–1984 (1990).
6. Broer, D. J., Lub, J. & Mol, G. N. Wide-band reflective polarizers from cholesteric polymer networks with a pitch gradient. *Nature* **378**, 467–469 (1995).
7. Coates, D. et al. New applications of liquid crystals and liquid crystal polymers. *SID 96 Applications Digest SID Int. Symp.* Vol. 27, 67–70 (Soc. for Information Display, Santa Ana, 1996).

8. Wortman, D. L. A recent advance in reflective polarizer technology. *SID 97 Applications Digest SID Int. Symp.* Vol. 28 (Information Display, Santa Ana, in the press).
9. Dirix, Y. *Polarizers Based on Anisotropic Absorbance or Scattering of Light*. Thesis, Tech. Univ. Eindhoven (1997).
10. Hagler, T. W. et al. Highly ordered conjugated polymers in polyethylene: orientation by mesoepitaxy. *Polymer Commun.* **32**, 339–342 (1991).
11. Weder, C., Sarwa, C., Bastiaansen, C. & Smith, P. Highly polarized luminescence from oriented conjugated polymer/polyethylene blend films. *Adv. Mater.* **9**, 1035–1039 (1997).
12. Weder, C., Sarwa, C., Montali, A., Bastiaansen, C. & Smith, P. New photoluminescent display devices. *Science* **279**, 835–837 (1998).
13. Weder, C., Sarwa, C., Bastiaansen, C. & Smith, P. Photoluminescent display devices. European Patent Application No. 97111229.7 (1997).
14. Tessler, N., Denton, G. J. & Friend, R. Lasing from conjugated-polymer microcavities. *Nature* **382**, 695–697 (1996).
15. Weder, C. & Wrighton, M. S. Efficient solid-state photoluminescence in new poly(2,5-dialkoxy-p-phenyleneethynylene)s. *Macromolecules* **29**, 5157–5165 (1996).
16. Guillet, J. *Polymer Photophysics and Photochemistry* (Cambridge Univ. Press, New York, 1985).
17. Webster, S. E. Photon-harvesting polymers. *Chem. Rev.* **90**, 1469–1482 (1990).
18. Vekshin, N. L. *Energy Transfer in Macromolecules* (SPIE Optical Engineering Press, Washington, 1997).
19. Gilbert, A. & Baggot, J. *Essentials of Molecular Photochemistry* (Blackwell Science, Cambridge, 1997).
20. Förster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.* **2**, 55–75 (1948).
21. Dexter, D. L. A theory of sensitized luminescence in solids. *J. Chem. Phys.* **21**, 836–850 (1953).
22. Knox, R. S. *Theory of Excitons* (Academic, New York, 1963).
23. Wudl, F. & Srdanov, G. Conducting polymer formed of poly(2-methoxy-5-(2'-ethyl-hexyloxy)-p-phenylenevinylene). US Patent No. 5189136 (1993).
24. Weder, C., Bastiaansen, C., Montali, A. & Smith, P. Efficient photoluminescent polarizers, process for forming and application in devices. European Patent Application No. 98101520.9 (1998).

Acknowledgements. We thank Ir. J. Visjager for help in preparing the manuscript.

Correspondence and request for materials should be addressed to C.W. (e-mail: weder@ifp.mat.ethz.ch).

Total synthesis of brevetoxin A

K. C. Nicolaou*, Zhen Yang, Guo-qiang Shi, Janet L. Gunzner, Konstantinos A. Agrios & Peter Gärtner

Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA and Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA

Brevetoxin A is the most potent neurotoxin secreted by *Gymnodinium breve* Davis, a marine organism often associated with harmful algal blooms known as 'red tides'^{1–3}. The compound, whose mechanism of action involves binding to and opening of sodium channels^{4–7}, is sufficiently toxic to kill fish at concentrations of nanograms per ml (refs 3, 4) and, after accumulation in filter-feeding shellfish, to poison human consumers. The precise pathway by which nature constructs brevetoxin A is at present unknown^{8,9}, but strategies for its total synthesis have been contemplated for some time. The synthetic challenge posed by brevetoxin A reflects the high complexity of its molecular structure: 10 oxygen atoms and a chain of 44 carbon atoms are woven into a polycyclic macromolecule that includes 10 rings (containing between 5 and 9 atoms) and 22 stereogenic centres. Particularly challenging are the 7-, 8- and 9-membered rings which allow the molecule to undergo slow conformational changes and force a 90° twist at one of its rings^{1–6}. Here we describe the successful incorporation of methods that were specifically developed for the construction of these rings^{10,11} into an overall strategy for the total synthesis of brevetoxin A in its naturally occurring form. The convergent synthesis reported here renders this scarce neurotoxin synthetically available and, more importantly, allows the design and synthesis of analogues for further biochemical studies.

Figure 1 shows the synthetic strategy for the construction of brevetoxin A (1, Fig. 1a) in bond disconnection (Fig. 1b) and retrosynthetic (Fig. 1c) formats. This analysis led to the expectation of using D-glucose (4, Fig. 1c) and D-mannose (5, Fig. 1c) as starting materials for the construction of the requisite BCDE (2, Fig. 1c) and GHIJ (3, Fig. 1c) advanced intermediates, respectively. The union of

2 and 3 under Horner–Wittig conditions was expected to furnish, in a highly convergent manner and after ring closure, the basic ring skeleton of the target molecule, from which brevetoxin A could be fashioned through deprotections and functional-group manipulations. At the outset of this project, it was also evident from this analysis that although the 5- and 6-membered rings of the target molecule could be derived via well-developed synthetic methods¹², the three larger rings were not easily accessible through conventional techniques. Thus, to construct the 7-, 8- and 9-membered rings we used two important reactions, namely the silver-promoted hydroxydithioketal cyclization reaction¹⁰ (rings F and G) and the palladium-catalysed functionalization of lactones via coupling of their cyclic ketene acetal phosphates¹¹ (rings B, D and E), that were specifically developed for this purpose.

The synthesis of the BCDE ring system 2 (Figs 1c and 4) of brevetoxin A began with D-glucose (4, Fig. 2) and proceeded via intermediates 16 (Fig. 2) and 26 (Fig. 3). Thus, the D-glucose-derived bis(acetonide) 6¹³ (Fig. 2) was selectively cleaved with H₅IO₆ (ref. 14; for abbreviations see legends in figures) to afford the corresponding aldehyde, which reacted with MeMgBr leading to the expected secondary alcohol (84% overall yield). Swern oxidation of this alcohol, followed by reaction with allylmagnesium bromide in the presence of Ti(*i*-PrO)₄, led to the stereoselective formation of tertiary alcohol 7 (76.5% overall yield from 6). Cleavage of the second acetonide moiety in 7 with EtSH–ZnCl₂ led to an open-chain thioketal, whose dibenzylation with NaH–BnBr and ring closure in the presence of I₂–aq. NaHCO₃ led smoothly to the 6-membered ring lactol 8 in 64.5% overall yield. Olefination of 8 with the appropriate stabilized Wittig reagent, followed by exposure to CSA furnished, stereoselectively, the tetrahydropyran system 9 (81% for two steps). Condensation of ketone 9 with CH₂ = C(OMe)OTBS (ref. 15) in the presence of ZnBr₂ gave

compound 10 as a mixture of diastereoisomers (inconsequential) in 93% yield. The terminal olefin in 10 was then converted to a methyl ketone via an oxymercuration–palladation procedure¹⁶ (90% yield), and thence to a vinyl triflate by reaction with NaHMDS–PhNTf₂ furnishing intermediate 11 (93% yield). The latter compound (11) was coupled¹⁷ with the zinc reagent IZnCH₂CH₂CO₂Me in the presence of Pd(Ph₃P)₄ catalyst to extend the ‘left-hand’ side-chain (85% yield). The product was then sequentially deprotected (LiOH; then Li in liquid NH₃) to afford the dihydroxydicarboxylic acid 12. Yamaguchi¹⁸ bis(lactonization) of 12 allowed the formation of bis(lactone) 13 (78.5% yield for three steps). Desilylation of 13 with HF–pyridine, followed by dehydration in the presence of

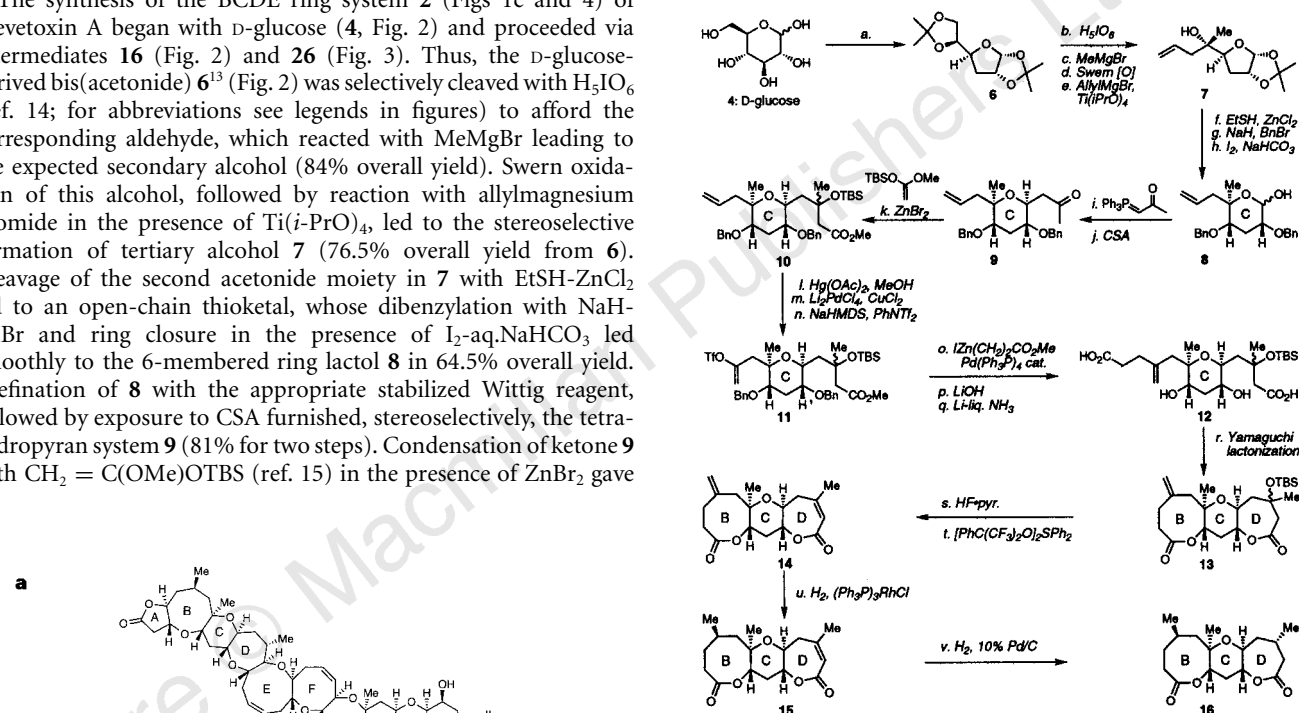


Figure 2 Construction of the BCD bis-lactone system 16. Reagents and conditions as follows. (a) See ref. 13; (b) H₅IO₆ (1.2 equiv.), EtOAc, 25 °C, 1 h; (c) MeMgBr (5.0 equiv.), Et₂O, 0 °C, 2 h, 84% for two steps; (d) oxalyl chloride (1.8 equiv.), DMSO (2.2 equiv.), Et₃N (5.0 equiv.), CH₂Cl₂, –78 °C, 1 h, 85%; (e) AllylMgBr (1.5 equiv.), Ti(*i*-PrO)₄ (1.5 equiv.), THF, –78 °C, 2 h, 90%; (f) EtSH (20.0 equiv.), ZnCl₂ (5.0 equiv.), CH₂Cl₂, 0 °C, 1 h, 94%; (g) NaH (3.0 equiv.), BnBr (2.0 equiv.), *n*-Bu₄N⁺PF₆[–] (cat), THF, 0 °C, 5 h, 80%; (h) I₂ (3.3 equiv.), NaHCO₃ (6.5 equiv.), acetone, 25 °C, 1 h, 86%; (i) Ph₃P = CHCOMe (1.5 equiv.), toluene, 110 °C, 4 h; (j) CSA (0.1 equiv.), CHCl₃, 25 °C, 1 h, 81% for two steps; (k) CH₂ = C(OMe)OTBS (1.5 equiv.), ZnBr₂ (0.5 equiv.), Et₂O, –78 °C, 2 h, 93%; (l) Hg(OAc)₂ (1.1 equiv.), MeOH, 25 °C, 2 h; (m) PdCl₂ (0.1 equiv.), LiCl (0.2 equiv.), CuCl₂ (3.0 equiv.), MeOH, 70 °C, 2 h, 90% for two steps; (n) NaHMDS (1.2 equiv.), PhNTf₂ (1.2 equiv.), THF, –78 °C, 1 h, 93%; (o) IZn(CH₂)₂CO₂Me (2.0 equiv.), Pd(Ph₃P)₄ (0.05 equiv.), DMA:benzene (1:10), 25 °C, 2 h, 85%; (p) LiOH (10.0 equiv.), THF:H₂O:MeOH (3:1:1), 65 °C, 3 h, 96%; (q) Li (12.0 equiv.), NH₃ (liq.), EtOH, –78 °C, 15 min; (r) 2,4,6-trichlorobenzoyl chloride (2.2 equiv.), Et₃N (4.0 equiv.), THF, 0 °C, 30 min; then 4-DMAP (6.0 equiv.), benzene, 75 °C, 4 h, 82% for two steps; (s) HF–pyridine (3.0 equiv.), CH₂Cl₂, 0 °C, 3 h; (t) [PhC(CF₃)₂O]₂SPh₂ (1.5 equiv.), CH₂Cl₂, 0 °C, 15 min, 81% for two steps; (u) H₂, (Ph₃P)₃RhCl (0.1 equiv.), benzene, 25 °C, 5 h, 99%, ~4:1 ratio; (v) H₂, 10% Pd/C (0.1 equiv.), EtOAc, 25 °C, 12 h, 95%, ~19:1 ratio; CSA, 10-camphorsulphonic acid; DMA, *N,N*-dimethylacetamide; 4-DMAP, 4-*N*-dimethylaminopyridine; DMF, *N,N*-dimethylformamide; DMSO, dimethylsulphoxide; NMO, 4-methylmorpholine-*N*-oxide; TBS, *t*-BuMe₂Si; THF, tetrahydrofuran; Tf, triflate. NaHMDS, sodium hexamethyldisilazide; PhNTf₂, *N*-phenyltrifluoromethanesulphonimide. For selected physical data for compound 14, see Supplementary Information.

Figure 1 Molecular structure (a), strategic bond disconnections (b), and retrosynthetic analysis (c) of brevetoxin A (1). Abbreviations for chemical groups: Me, methyl; Ph, phenyl; Tr, trityl; TBDPS, *t*-butyldiphenylsilyl; Et, ethyl; TBS, *t*-butyldimethylsilyl.

Martin's sulphurane ($[\text{PhC}(\text{CF}_3)_2\text{O}]_2\text{SPh}_2$) (ref. 19), furnished the doubly unsaturated crystalline bis(lactone) **14** in 81% overall yield. (The melting point of **14** was 158–159 °C (ether/hexane)). An X-ray crystallographic analysis confirmed the structure of **14**. Sequential hydrogenation of the two olefinic bonds in **14**, first with Wilkinson's catalyst $[(\text{Ph}_3\text{P})_3\text{RhCl}]$ (**15**, 99% yield, ~4:1 ratio)²⁰ and then with Pd/C (**16**, 95% yield, ~19:1 ratio), established the two methyl groups in their correct stereochemical arrangement. Compound **16** was separated from its minor diastereoisomers by silica-gel flash chromatography.

The conversion of bis(lactone) **16** to the tetracyclic 9-membered ring lactone **26** relied on a two-directional approach similar to that previously reported from these laboratories²¹ and on a newly developed method for the transformation of lactones to cyclic ethers involving cyclic ketene acetal phosphates¹¹. Figure 3 summarizes the sequence for this conversion. The bis(lactone) **16** was added to a solution of KHMDS and $(\text{PhO})_2\text{P}(\text{O})\text{Cl}$, furnishing bis(phosphate) **17** (80% yield), which was coupled with $\text{Me}_3\text{SnSnMe}_3$ in the presence of $\text{Pd}(\text{Ph}_3\text{P})_4$ catalyst and LiCl to afford the bis(vinylstannane) **18** (81% yield). Reaction of **18**

with *n*-BuLi, followed by addition of *n*-PrC $\equiv$ CCu-2HMPT (ref. 22) and $\text{TrOCH}_2\text{CH}_2\text{OBn}$, led to bis(vinylether) **20** in 65% overall yield via mixed cuprate **19**. Hydroboration of **20** with the xylborane, followed by basic hydrogen peroxide work-up (80% yield) and selective monosilylation (90% yield, five recycles) as previously described²¹, led to compound **22** via **21**. Having simultaneously functionalized bis(lactone) **16** on both sides, and then successfully differentiating the two ends, we were now in a position to resume the stepwise manipulation of the molecule towards the 9-membered ring lactone **26**. Thus, hydrogenolysis of the two benzyl ethers in **22** (H_2 , 10% Pd/C, 93% yield) and ketalization with cyclohexanone dimethylketal in the presence of PPTS (85% yield), followed by pivalate formation (PivCl, 4-DMAP, 99% yield) at the remaining primary alcohol, gave compound **23**. The ketal was then disassembled under acidic conditions (CSA, 88% yield), and the two hydroxyl groups so released were differentiated by trityl ether formation (TrCl-4-DMAP, primary alcohol, 95% yield), acetylation (Ac_2O , 4-DMAP, secondary alcohol, 99% yield) and trityl ether cleavage (TFA, 90% yield) to afford hydroxy acetate **24**. The 9-membered ring lactone was then built on the 'right side' of the

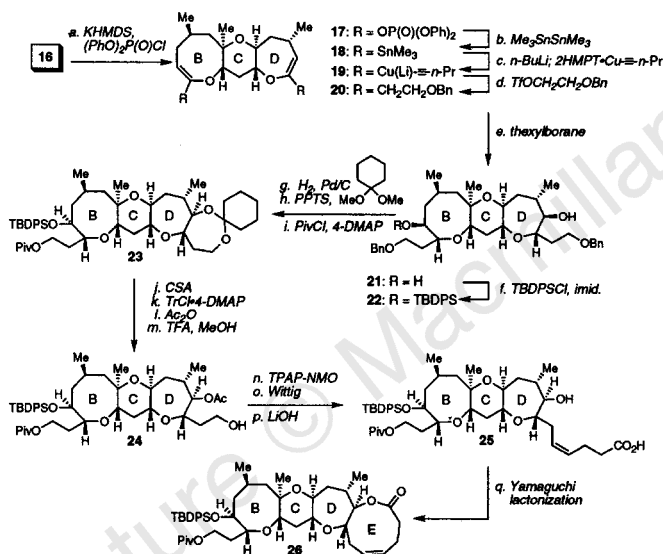


Figure 3 Construction of the BCDE lactone **26**. Reagents and conditions as follows. (a) KHMDS (4.0 equiv.), $(\text{PhO})_2\text{P}(\text{O})\text{Cl}$ (6.0 equiv.), HMPA (6.0 equiv.), THF, –78 °C, 1 h, 80%; (b) $\text{Me}_3\text{SnSnMe}_3$ (8.0 equiv.), LiCl (6.0 equiv.), $\text{Pd}(\text{Ph}_3\text{P})_4$ (0.1 equiv.), THF, 70 °C, 2 h, 81%; (c) *n*-BuLi (1.55 M in hexanes, 4.0 equiv.), $\text{Cu}-\text{C}\equiv\text{C}-n\text{-Pr}$ (4.4 equiv.), HMPT (8.8 equiv.), THF, –78 → –30 °C, 2 h; (d) $\text{TrOCH}_2\text{CH}_2\text{OBn}$ (7.0 equiv.), THF, –78 → –30 °C, 12 h, 65% for two steps; (e) the xylborane (4.0 equiv.), THF, –30 → 0 °C, 3 h; then H_2O_2 (20.0 equiv.), NaOH (20.0 equiv.), H_2O , 0 °C, 2 h, 80%; (f) *t*-BuPh₂SiCl (1.3 equiv.), imidazole (3.0 equiv.), DMF, 25 °C, 8 h, 90% (five recycles); (g) 10% Pd/C (10% wt. equiv.), H_2 , MeOH, 25 °C, 2 h, 93%; (h) cyclohexanone dimethylketal (1.3 equiv.), PPTS (0.1 equiv.), CH_2Cl_2 , 25 °C, 2 h, 85%; (i) PivCl (1.3 equiv.), 4-DMAP (1.4 equiv.), CH_2Cl_2 , 25 °C, 0.5 h, 99%; (j) CSA (0.1 equiv.), MeOH: CH_2Cl_2 (2:1), 25 °C, 0.5 h, 88%; (k) TrCl-4-DMAP (4.0 equiv.), CH_2Cl_2 , 40 °C, 12 h, 95%; (l) Ac_2O (1.5 equiv.), 4-DMAP (0.5 equiv.), CH_2Cl_2 , 25 °C, 15 min, 99%; (m) TFA (1.5 equiv.), MeOH (10 equiv.), CH_2Cl_2 , 25 °C, 30 min, 90%; (n) TPAP (0.1 equiv.), NMO (2.0 equiv.), CH_2Cl_2 , 25 °C, 30 min, 97%; (o) $^{\text{Br}}\text{PPh}_3(\text{CH}_2)_3\text{CO}_2\text{Me}$ (2.5 equiv.), KHMDS (2.0 equiv.), THF, –78 °C, 1 h, 83%; (p) LiOH (10.0 equiv.), THF: H_2O :MeOH (3:2:1), 0 °C, 4 h; then 4-DMAP (4.0 equiv.), benzene, 80 °C, 2 h, 89% for two steps. 4-DMAP, 4-*N*-dimethylaminopyridine; DMF, *N,N*-dimethylformamide; KHMDS, potassium hexamethyldisilazide; HMPT, hexamethylphosphorous triamide; NMO, 4-methylmorpholine-*N*-oxide; Piv, pivalate; PPTS, pyridinium *p*-toluenesulphonate; TPAP, tetra-*n*-propylammonium perruthenate; TBDPS, *t*-BuPh₂Si; Tr, trityl. For selected physical data for compound **26**, see Supplementary Information.

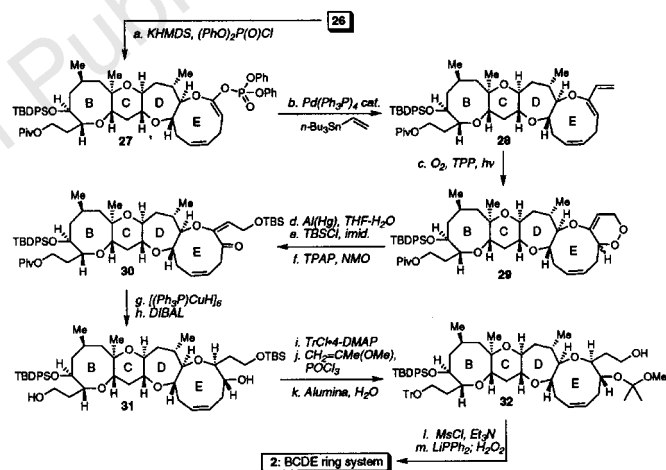


Figure 4 Construction of the BCDE phosphine oxide **2**. Reagents and conditions as follows. (a) KHMDS (3.0 equiv.), $(\text{PhO})_2\text{P}(\text{O})\text{Cl}$ (5.0 equiv.), THF, –78 °C, 1 h; (b) (*n*-Bu)₃SnCH=CH₂ (3.0 equiv.), LiCl (5.0 equiv.), $\text{Pd}(\text{Ph}_3\text{P})_4$ (0.1 equiv.), THF, 75 °C, 2 h, 81% for two steps; (c) O_2 , TPP (0.01 equiv.), CCl_4 , 25 °C, 0.3 h; (d) Al(Hg) (excess), H_2O :THF (1:29), 25 °C, 2 h, 58% for two steps; (e) *t*-Bu₂SiCl (1.5 equiv.), imidazole (10.0 equiv.), CH_2Cl_2 , 25 °C, 1 h, 91%; (f) TPAP (0.1 equiv.), NMO (2.0 equiv.), 4-Å MS, CH_2Cl_2 , 25 °C, 1 h, 82%; (g) $[(\text{Ph}_3\text{P})\text{CuH}]_6$ (2.0 equiv.), benzene, 25 °C, 72 h, 70%; (h) DIBAL (2.5 equiv.), CH_2Cl_2 , –78 °C, 30 min, 95%, ~5:1 ratio; (i) TrCl-4-DMAP (15.0 equiv.), CH_2Cl_2 , 40 °C, 24 h, 75% pure trityl ether after silica-gel chromatography; (j) POCl_3 (0.01 equiv.), $\text{CH}_2=\text{CMe}(\text{OMe})$ (solvent), 25 °C, 6 h, 95%; (k) neutral alumina-1% H_2O activated by heating (excess), hexane, 25 °C, 2 h, 96%; (l) MsCl (2.0 equiv.), Et₃N (4.0 equiv.), CH_2Cl_2 , 0 °C, 15 min, 99%; (m) Ph₂PLi (3.0 equiv.), HMPA (3.0 equiv.), THF, 0 °C, 30 min; then 5% aq. H_2O_2 (excess), 93%; DIBAL, diisobutylaluminum hydride; 4-DMAP, 4-*N*-dimethylaminopyridine; HMPA, hexamethylphosphoramide; KHMDS, potassium hexamethyldisilazide; MS, molecular sieves; Et₃N, 4-methylmorpholine-*N*-oxide; TBDPS, *t*-BuPh₂Si; TBS, *t*-BuMe₂Si; THF, tetrahydrofuran; TPAP, tetra-*n*-propylammonium perruthenate; TPP, tetraphenylporphyrin; Tr, trityl. Ms, CH_3SO_2 or methanesulphonyl; imid., imidazole. For selected physical data for compound **2**, see Supplementary Information.

molecule by (1) TPAP-NMO oxidation²³ of the primary alcohol to the corresponding aldehyde (97% yield); (2) olefination with the ylide derived from $\text{Br}^-\text{Ph}_3\text{P}^+(\text{CH}_2)_3\text{CO}_2\text{Me}$ and KHMDS (83% yield); (3) saponification of both the methylester and the acetate groups (LiOH) to afford hydroxyacid **25**; and (4) Yamaguchi¹⁸ lactonization, furnishing the targeted intermediate **26** (89% for two steps).

The functionalization of the 9-membered ring lactone to the desired E ring ether functionality was achieved via sequential application of the cyclic ketene acetal phosphate-palladium coupling technology¹¹ and a selective²⁴ singlet oxygen 4 + 2 cycloaddition reaction of a constructed conjugated diene system. Figure 4 outlines the chemistry leading from **26** to the targeted phosphine oxide **2**. Conversion of **26** to its ketene acetal phosphate by reaction with KHMDS and $(\text{PhO})_2\text{P}(\text{O})\text{Cl}$, followed by palladium-catalysed coupling with tri-*n*-butylvinyltin, resulted, via **27**, in the formation of conjugated diene **28** (81% overall yield), thus accomplishing the crucial transformation of a lactone to a cyclic ether¹¹. The selective functionalization of the diene system in **28** in the presence of the additional double bond within the 9-membered ring, was achieved

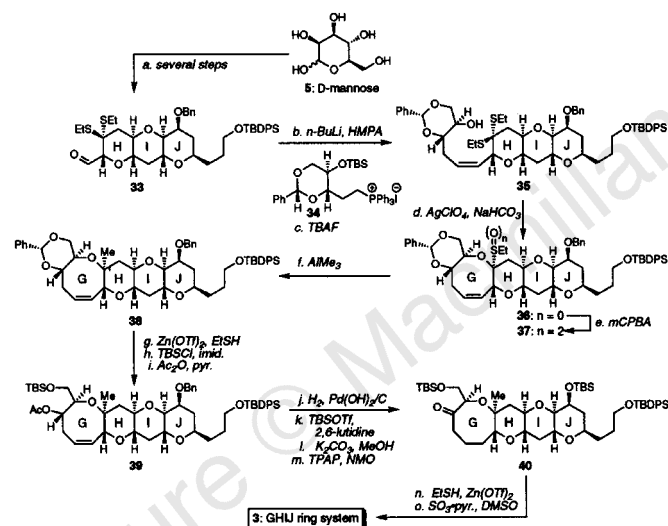


Figure 5 Construction of the GHJ ring system **3**. Reagents and conditions as follows. (a) See ref. 27; (b) add *n*-BuLi (1.55 M in hexanes, 1.0 equiv.) to **34** (1.0 equiv.) and HMPA (10.0 equiv.) in THF; then add **33**, THF, -78°C , 1 h; then 25°C , 1.5 h, 88%; (c) TBAF (1.1 equiv.), THF, 25°C , 7 h, 83%; (d) AgClO_4 (3.0 equiv.), NaHCO_3 (3.0 equiv.), SiO_2 , 4-Å MS, MeNO_2 , 25°C , 3 h, 92%; (e) *m*CPBA (2.0 equiv.), CH_2Cl_2 , 0°C , 2 h, 93%; (f) AlMe_3 (3.0 equiv.), CH_2Cl_2 , -78°C , 1 h, 94%; (g) EtSH (20 equiv.), $\text{Zn}(\text{OTf})_2$ (0.3 equiv.), NaHCO_3 (0.5 equiv.), CH_2Cl_2 , 25°C , 4 h, 92%; (h) TBSOTf (1.1 equiv.), imidazole (1.5 equiv.), CH_2Cl_2 , 25°C , 1 h, 92%; (i) Ac_2O (1.1 equiv.), 4-DMAP (1.3 equiv.), CH_2Cl_2 , 25°C , 2.5 h, 95%; (j) H_2 , 10% $\text{Pd}(\text{OH})_2/\text{C}$ (0.15 equiv.), AcOH , 25°C , 5 h, 90%; (k) TBSOTf (2.5 equiv.), 2,6-lutidine (3.0 equiv.), CH_2Cl_2 , 0°C , 0.5 h, 87%; (l) K_2CO_3 (0.5 equiv.), MeOH , 25°C , 4 h, 93%; (m) TPAP (0.05 equiv.), NMO (2.0 equiv.), CH_2Cl_2 , 25°C , 1 h, 94%; (n) EtSH (20.0 equiv.), $\text{Zn}(\text{OTf})_2$ (0.3 equiv.), CH_2Cl_2 , 25°C , 2.5 h, 92%; (o) SO_3 :pyridine (3.0 equiv.), DMSO (10.0 equiv.), Et_3N (10.0 equiv.), CH_2Cl_2 , 0°C , 1.5 h, 87%. 4-DMAP, 4-*N*-dimethylaminopyridine; DMF, *N,N*-dimethylformamide; DMSO, dimethyl sulphoxide; HMPA, hexamethylphosphoramide; *m*CPBA, *m*-chloroperoxybenzoic acid; NMO, 4-methylmorpholine-*N*-oxide; TBAF, tetra-*n*-butylammonium fluoride; TBDS, *t*-BuPh₂Si; TBS, *t*-BuMe₂Si; THF, tetrahydrofuran; TPAP, tetra-*n*-propylammonium perruthenate; Tr, trityl. Selected physical data for compound **3**, see Supplementary Information.

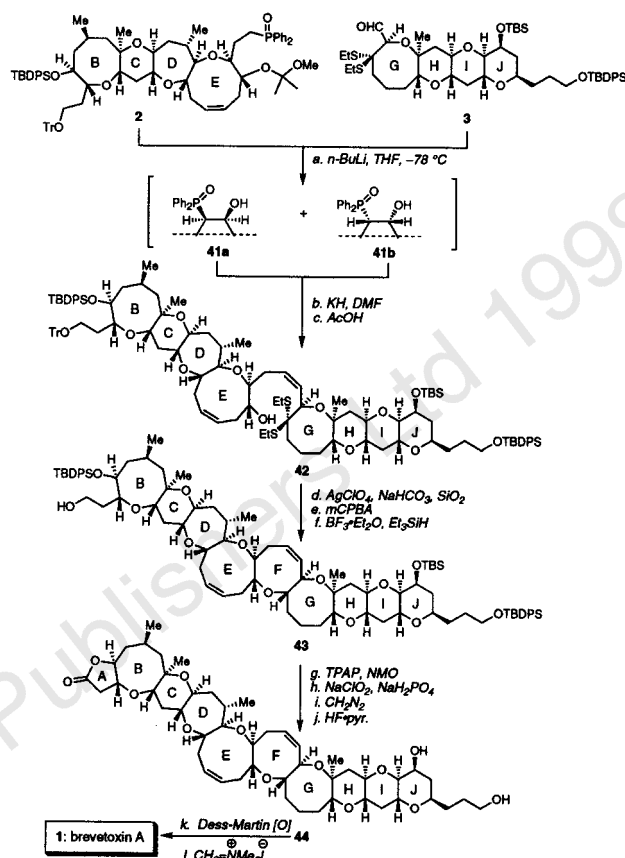


Figure 6 Total synthesis of brevetoxin A (**1**). Reagents and conditions as follows. (a) Add *n*-BuLi (1.55 M in hexanes, 1.0 equiv.) to **2** (1.0 equiv.) in THF at -78°C , then **3**, 20 min, 75%; (b) KH (2.0 equiv.), DMF, 25°C , 0.5 h; (c) 20% aq. AcOH, THF, 25°C , 6 h, 72% for two steps; (d) AgClO_4 (3.0 equiv.), NaHCO_3 (3.0 equiv.), SiO_2 , 4-Å MS, MeNO_2 , 25°C , 2 h, 80%; (e) *m*CPBA (2.2 equiv.), CH_2Cl_2 , 0°C , 1 h, 85%; (f) $\text{BF}_3\cdot\text{Et}_2\text{O}$ (3.0 equiv.), Et_3SiH , CH_2Cl_2 , -78°C , 30 min, 80%; (g) TPAP (0.1 equiv.), NMO (2.0 equiv.), CH_2Cl_2 , 25°C , 30 min; (h) NaClO_2 (2.0 equiv.), NaH_2PO_4 (3.0 equiv.), 2-methyl-2-butene (5.0 equiv.), *t*-BuOH:H₂O (5:1), 25°C , 30 min; (i) CH_2N_2 , Et_2O , 25°C , 15 min, 80% for three steps; (j) HF-pyridine (10.0 equiv.), CH_2Cl_2 , 0°C , 2 h, 90%; (k) Dess Martin periodinane (1.5 equiv.), CH_2Cl_2 , 25°C , 1 h, 80%; (l) $\text{CH}_2 = \text{N}^+\text{Me}_2\text{I}^-$ (10.0 equiv.), Et_3N (15.0 equiv.), CH_2Cl_2 , 25°C , 5 h, 90%. DMF, *N,N*-dimethylformamide; *m*CPBA, *m*-chloroperoxybenzoic acid; NMO, 4-methylmorpholine-*N*-oxide; TBDS, *t*-BuPh₂Si; TBS, *t*-BuMe₂Si; THF, tetrahydrofuran; TPAP, tetra-*n*-propylammonium perruthenate; Tr, trityl. Selected physical data for synthetic brevetoxin A (**1**): retention factor $R_f = 0.45$ (silica, 1:1 ethyl acetate:hexane); specific rotation $[\alpha]_D^{25} + 75.5$ (c 0.2, CH_2Cl_2); infrared spectra (pure compound) ν_{max} 3,425.1, 2,920.5, 2,849.2, 1,788.2, 1,689.5, 1,455.1, 1,378.7, 1,314.1, 1,265.7, 1,208.3, 1,079.9, 897.1, 737.0 cm^{-1} ; ^1H NMR (C_6D_6 , 600 MHz, 45°C) δ 9.29 (s, 1H), 5.90 (s, 1H), 5.87 (dd, $J = 10.8, 5.2$ Hz, 1H), 5.78–5.67 (bm, 2H), 5.71–5.64 (m, 1H), 5.43 (s, 1H), 4.55 (dd, $J = 8.0, 5.4$ Hz, 1H), 3.98–3.91 (m, 3H), 3.74 (b, 1H), 3.59–3.54 (m, 1H), 3.52 (ddd, $J = 10.6, 7.9, 2.3$ Hz, 1H), 3.46–3.34 (m, 4H), 3.29 (bd, $J = 8.5$ Hz, 1H), 3.19–3.10 (m, 3H), 2.94 (ddd, $J = 12.0, 8.0, 4.1$ Hz, 1H), 2.89 (ddd, $J = 11.9, 9.2, 4.5$ Hz, 1H), 2.80 (dd, $J = 11.9, 4.6$ Hz, 1H), 2.78 (dd, $J = 9.6, 2.7$ Hz, 1H), 2.72–2.40 (bm, 4H), 2.65 (dd, $J = 10.7, 10.7$ Hz, 1H), 2.40 (ddd, $J = 15.2, 10.8, 4.0$ Hz, 1H), 2.36 (dd, $J = 16.9, 8.2$ Hz, 1H), 2.31 (dd, $J = 11.6, 4.2$ Hz, 1H), 2.25–2.21 (m, 1H), 2.24–2.20 (m, 1H), 2.21 (dd, $J = 16.9, 10.3$ Hz, 1H), 2.18–2.10 (m, 1H), 2.05–1.92 (m, 6H), 1.92–1.86 (m, 1H), 1.86–1.74 (m, 3H), 1.72–1.52 (m, 7H), 1.44–1.38 (m, 1H), 1.37–1.24 (m, 1H), 1.24 (s, 3H), 1.20 (d, $J = 7.1$ Hz, 3H), 0.98–0.95 (m, 1H), 0.96 (s, 3H), 0.77 (d, $J = 6.8$ Hz, 3H); ^{13}C NMR (C_6D_6 , 150 MHz, 45°C) δ 193.5, 171.0, 148.9, 139.0, 134.4, 128.5, 127.4, 124.8, 92.0, 88.5, 86.7, 85.1, 84.7, 84.1, 83.8, 82.2, 82.0, 80.4, 79.9, 78.9, 77.8, 76.7, 76.2, 71.9, 71.1, 69.6, 66.3, 62.4, 52.9, 45.6, 43.3, 38.2, 37.5, 36.9, 36.2, 35.9, 33.8, 32.3, 31.4, 29.0, 27.6, 27.2, 21.4, 19.7, 16.8, 15.0 (carbons 17, 20 and 28 were observed at best as weak broad signals at 34.5, 34.5 and 30.9, respectively, as previously reported^{30,31}). Mass spectra; calculated for $\text{C}_{49}\text{H}_{70}\text{O}_{13}$ ($M + \text{Na}^+$) 889.4714, found 889.4747. For selected physical data for compounds **42** and **43**, see Supplementary Information.

by a novel 4 + 2 cycloaddition reaction²⁴ involving singlet O₂, generated by light in the presence of tetraphenylporphyrin (TPP) as a sensitizer. The formed endoperoxide **29** (~1:1 ratio of diastereoisomers, inconsequential) was then converted to the corresponding diol (58% yield for two steps) by reductive cleavage of the O–O bond with Al(Hg) in THF–H₂O. The primary hydroxyl group of the resulting diol was selectively silylated with TBSCl-imid. (91% yield), and the secondary alcohol was oxidized with TPAP-NMO (ref. 23; 82% yield), leading to enone **30**.

The proper functionality and stereochemistry was established on ring E by sequential reduction of the double bond ([Ph₃P]CuH]₆, 70% yield)²⁵ and of the carbonyl group (DIBAL). The DIBAL reduction also cleaved the pivalate group, furnishing diol **31** (95% yield, ~5:1 ratio of diastereoisomers, separated at a later stage). The resulting diol (**31**) was protected by sequential formation of a trityl ether at the primary position (TrCl·4-DMAP, separation of diastereoisomers with silica-gel flash chromatography, 75% pure trityl ether) and of a mixed methoxyketal at the secondary position [CH₂ = CMe(OMe), POCl₃ catalyst], before selective removal of the TBS group (alumina)²⁶ to afford primary alcohol **32** (91% yield for two steps). Finally, mesylation of **32** (MsCl, Et₃N), followed by displacement with LiPPh₂ and oxidation with hydrogen peroxide, furnished the desired phosphine oxide **2** (BCDE ring system) in 92% overall yield.

The construction of the other requisite advanced intermediate, GHII ring system **3**, is shown in Fig. 5. Thus, D-mannose (**5**) was converted to tricyclic dithioketal-aldehyde **33** by a modification of our previously published²⁷ multi-step sequence (full details will be published elsewhere). Generation of the ylide from phosphonium salt **34**²⁷ (*n*-BuLi) and coupling with aldehyde **33** (THF, HMPA, –78 °C) resulted, after selective removal of the TBS group, in the stereoselective formation of *Z*-olefin **35** in 73% overall yield. Hydroxydithioketal cyclization¹⁰ of **35**, facilitated by AgClO₄ in the presence of NaHCO₃, SiO₂ 4-Å MS in MeNO₂, resulted in the formation of the mixed 8-membered ring thioketal **36** in 92% yield. The introduction of the α-methyl group at the fusion of rings G and H was accomplished by oxidation of **36** to sulphone **37** (*m*CPBA, 93% yield) followed by reaction with AlMe₃ to afford **38** (94% yield). The well-defined conformation of the intermediate oxonium species in the latter reaction is presumed to be responsible for the observed stereochemical control. The benzylidene group in **38** was then removed by exposure to EtSH in the presence of Zn(OTf)₂ and NaHCO₃ (92% yield), and the resulting diol was differentiated by silylation (TBSCl-imid., primary silyl ether, 92% yield), followed by acetylation (Ac₂O, pyridine, secondary acetate, 95% yield), leading to compound **39**. Hydrogenation of the double bond in **39** in the presence of Pd(OH)₂/C resulted in concomitant debenzoylation, furnishing the corresponding saturated alcohol in 90% yield. The newly liberated secondary hydroxyl group was then protected as a TBS ether (TBSOTf-2,6-lutidine, 87% yield), and the acetate group on ring G was cleaved by the action of K₂CO₃ in MeOH (93% yield). The resulting secondary alcohol was then oxidized to ketone **40** with TPAP-NMO (ref. 23) in 94% yield. Finally, exposure of **40** to EtSH in the presence of Zn(OTf)₂ led to the corresponding dithioketal with concomitant desilylation of the neighbouring oxygen (92% yield), and was followed by oxidation (SO₃·pyridine, DMSO, Et₃N)²⁸ to afford the targeted GHII ring system **3**, in 87% overall yield.

The union of fragments **2** and **3** and the final stages of the total synthesis of brevetoxin A (**1**) are depicted in Fig. 6. Thus, generation of the anion of phosphine oxide **2** with *n*-BuLi (THF, –78 °C), followed by addition of aldehyde **3**, gave a mixture of two diastereomeric hydroxyphosphine oxides (**41a** and **41b**, ~1:1 ratio, 75% total yield). Treatment of this mixture (**41a** + **41b**) with KH in DMF, followed by acidic cleavage of the mixed ketal, resulted in the formation of the desired *Z*-olefin **42** as a single isomer and in 72% yield. The choice of the phosphine oxide (Horner–Wittig

reaction)²⁹ as opposed to the more commonly used phosphonium salt for fragment **2** was necessitated by the inability of the latter to react, via its ylide, with the sterically hindered aldehyde **3**. Furthermore, the placement of the mixed methoxyketal on ring E of the fragment **2** was crucial for the exclusive *Z*-stereoselectivity observed in the coupling reaction. It is presumed that the two oxygens of this group act synergistically with the E-ring oxygen in complexing the lithium atom in a well-defined structure, which leads to the exclusive formation of only the *anti* adducts **41a** and **41b**. Both diastereoisomers furnish the same *Z*-olefin (whereas the *syn* diastereoisomers would form the *E*-olefin). Ring closure of hydroxydithioketal¹⁰ **42**, induced by AgClO₄, furnished the 8-membered ring mixed thioketal in 80% yield. Oxidation of this mixed thioketal with *m*CPBA led to the corresponding sulphone (85% yield), which upon exposure to BF₃·Et₂O in the presence of Et₃SiH (ref. 10) at –78 °C, resulted in the reductive removal of the ethyl sulphone moiety and concomitant cleavage of the trityl ether, furnishing the BCDEFGHII ring skeleton (compound **43**, 80% yield) of brevetoxin A. The stereochemistry of the newly generated stereocenter at the FG ring junction of **43** was confirmed by comparison of its NMR data to those of brevetoxin A^{1,3,30,31} and of a model system²⁴. The fusion of ring A onto the rest of the brevetoxin A framework was accomplished by stepwise oxidation of the primary alcohol in **43** to the corresponding aldehyde (TPAP-NMO)²³ and carboxylic acid (NaClO₂–NaH₂PO₄, 2-methyl-2-butene)³², followed by methyl ester formation (CH₃N₂, 80% for three steps) and exposure to HF-pyridine. The latter reaction resulted in the removal of all three silyl groups and spontaneous ring closure to form the γ-lactone ring, furnishing the ABCDEFGHII ring system **44** (90% yield) of brevetoxin A. Finally, exposure of diol **44** to carefully controlled amounts of Dess–Martin³³ reagent gave, selectively, the corresponding hydroxy aldehyde (80% yield), whose reaction with CH₂ = N⁺Me₂I[–] (ref. 34) proceeded smoothly to afford brevetoxin A (**1**) in 90% yield. Synthetic brevetoxin A was identical in all respects with naturally occurring brevetoxin A (as determined by thin-layer chromatography, high-performance liquid chromatography, ¹H and ¹³C NMR, optical rotation [α]_D²², infrared spectroscopy and mass spectrometry)^{1–3,30,31}.

The total synthesis of brevetoxin A (**1**) became possible only after a long campaign which involved several thwarted strategies and the development of a number of new synthetic technologies. Particularly rewarding were the solutions found to the synthetic challenges posed by the 7-, 8- and 9-membered rings of the target molecule^{10,11} and the convergent strategy finally developed to reach the complex and aesthetically pleasing structure of brevetoxin A. Through the described chemistry, brevetoxin A (**1**) joins brevetoxin B^{35–38} in yielding to total synthesis. The work described here opens up the possibility of preparing suitable molecular probes that may contribute to the development of sensing devices³⁹ for this and related biotoxins of the ‘red tides’^{40–45} and enhance our understanding of this phenomenon. □

Received 24 November 1997; accepted 19 January 1998.

1. Shimizu, Y., Chou, H.-N. & Bando, H. Structure of brevetoxin A (GB-1 toxin), the most potent toxin in the Florida red tide organism *Gymnodinium breve* (*Pytychodiscus brevis*). *J. Am. Chem. Soc.* **108**, 514–515 (1986).
2. Shimizu, Y., Bando, H., Chou, H.-N., Van Duyne, G. & Clardy, J. C. Absolute configuration of brevetoxins. *J. Chem. Soc., Chem. Commun.* 1656–1658 (1986).
3. Pawlak, J. *et al.* Structure of brevetoxin A as constructed from NMR and MS data. *J. Am. Chem. Soc.* **109**, 1144–1150 (1987).
4. Rein, K. S., Baden, D. G. & Gawley, R. E. Conformational analysis of the sodium channel modulator, brevetoxin A, comparison with brevetoxin B conformations, and a hypothesis about the common pharmacophore of the ‘site 5’ toxins. *J. Org. Chem.* **59**, 2101–2106 (1994).
5. Rein, K. S., Lynn, B., Baden, D. G. & Gawley, R. E. Brevetoxin B: chemical modifications, synaptosome binding, toxicity, and an unexpected conformational effect. *J. Org. Chem.* **59**, 2107–21013 (1994).
6. Yasumoto, T. & Murata, M. Marine toxins. *Chem. Rev.* **93**, 1897–1909 (1993).
7. Gawley, R. E. *et al.* The relationship of brevetoxin ‘length’ and A-ring functionality to binding and activity in neuronal sodium channels. *Chem. Biol.* **2**, 533–541 (1995).
8. Chou, H.-N. & Shimizu, Y. Biosynthesis of brevetoxins. Evidence for the mixed origin of the backbone carbon chain and the possible involvement of dicarboxylic acids. *J. Am. Chem. Soc.* **109**, 2184–2185 (1987).
9. Lee, M. S., Qin, G.-W., Nakanishi, K. & Zagorski, M. G. Biosynthesis studies on brevetoxins, potent

An unusual mid-Pleistocene monsoon period over Africa and Asia

M. Rossignol-Strick*, M. Paterne†, F. C. Bassinot†, K.-C. Emeis‡ & G. J. De Lange§

* Laboratoire de Paléontologie et Palynologie, Box 106, Université Pierre et Marie Curie, 75005 Paris, France

† Centre des Faibles Radioactivités, Laboratoire mixte CNRS-CEA, Domaine du CNRS, avenue de la Terrasse, 91198 Gif-sur-Yvette, France

‡ Institut für Ostseeforschung Warnemuende, Seestr. 15, 18119 Warnemuende, Germany

§ Department of Geochemistry, Institute of Earth Sciences, Utrecht University, Budapestlaan 4, 3584 CD Utrecht, Netherlands

During the Quaternary period, organic-rich black layers called sapropels were intermittently deposited in the deep eastern Mediterranean Sea^{1,2} following high flood periods of the Nile River³. During the past 250 kyr, timing of sapropel formation coincides with astronomically driven maximum summer insolation in the northern tropics⁴. The insolation variations—described by a monsoon index⁴—modulate the intensity of the African monsoon feeding the Nile flood⁴. Here, we report the observation of a thick sapropel in eastern Mediterranean sediments that conspicuously deviates from the usual pattern. The sapropel, dated at 528–525 kyr by astronomical tuning of the stratigraphic oxygen-isotopic record, is anomalous because the tropical summer insolation, while at a peak at this time, was much lower than during the deposition of the more recent sapropels. The Mediterranean climate was cold and dry, at least at sea level. At the same time, in the equatorial Indian Ocean there was an extreme event of low surface-water salinity caused by heavy monsoonal fluvial discharge⁵. The simultaneity, within dating uncertainties, of unusually heavy monsoon rainfall over Africa and Asia while summer insolation (and the monsoon index) was relatively low indicates a large, regional-scale monsoon anomaly that cannot be explained in terms of current understanding of astronomical forcing.

Figure 1 shows the high-resolution oxygen isotope record of the sediment core, from the Ionian Sea, in which the unusual sapropel was identified. The timescale has been developed by tuning the oxygen isotope record to the ice volume model⁶, based on the July insolation at 65° N (ref. 7 and shown in Fig. 1b). The age of the core bottom is estimated to be ~1,200 kyr, in agreement with palaeomagnetic investigations⁸. Below the eight youngest sapropels^{9,10}, only six are seen between 250 and 1,200 kyr, although more could be expected from their apparent dependence on insolation variation, expressed as monsoon index in Fig. 1c. Some may have formed but disappeared by oxidation after deposition^{11,12}. These six sapropels presumably were the thickest of the original series because the downward progression of the oxidation front did not reach the sapropel base. Figure 1a also shows some 'ghosts' of sapropels traced by barium enrichment and remanent magnetization^{8,13,28}. Nevertheless, the absence of traces of many expected sapropels—according to the monsoon index maxima—is noticeable.

Sapropel A is seen from 2,301 to 2,295 cm, and shows a very depleted $\delta^{18}\text{O}$ value, as in all the other sapropels. Figure 2a, b shows its detailed chronological and isotopic setting between interglacial stages 13 and 15. Its tuned-up age extends from 528 kyr at 2300.5 cm to 525 kyr at 2,295 cm. It coincides in the target ice-volume model with a weak deglacial peak of small amplitude at ~525 kyr at the end of interglacial stage 15. The astronomical solution of Berger and Loutre²⁹ is inherently more accurate for this time period³⁰, and yields an identical age. The monsoon index shown in Fig. 2e has a

- neurotoxins produced by the dinoflagellate *Gymnodinium breve*. *J. Am. Chem. Soc.* **111**, 6234–6241 (1989).
10. Nicolaou, K. C., Prasad, C. V. C., Hwang, C.-K., Duggan, M. E. & Veale, C. A. Cyclizations of hydroxydithioketals. New synthetic technology for the construction of oxocenes and related medium ring systems. *J. Am. Chem. Soc.* **111**, 5321–5330 (1989).
11. Nicolaou, K. C., Shi, G.-Q., Gunzner, J. L., Gärtner, P. & Yang, Z. Palladium-catalyzed functionalization of lactones via their cyclic ketene acetal phosphates. Efficient new synthetic technology for the construction of medium and large cyclic ethers. *J. Am. Chem. Soc.* **119**, 5467–5468 (1997).
12. Nicolaou, K. C., Prasad, C. V. C., Somers, P. K. & Hwang, C.-K. Activation of 6-endo over 5-exo hydroxy epoxide openings. Stereo- and ringselective synthesis of tetrahydrofuran and tetrahydropyran systems. *J. Am. Chem. Soc.* **111**, 5330–5334 (1989).
13. Freeman, J. P. in *Organic Syntheses (Collective Vol. VII)* 139–141 (Wiley, New York, 1990).
14. Xie, M., Beges, D. A. & Robins, M. J. Efficient "dehomologation" of di-O-isopropylidenehexofuranose derivatives to give O-isopropylidenehexofuranose by sequential treatment with periodic acid in ethyl acetate and sodium borohydride. *J. Org. Chem.* **61**, 5178–5179 (1996).
15. Mukaiyama, T., Banno, K. & Narasaka, K. New cross-aldol reactions. Reactions of silyl enol ethers with carbonyl compounds activated by titanium tetrachloride. *J. Am. Chem. Soc.* **96**, 7503–7509 (1974).
16. Rodeheaver, G. T. & Hunt, D. F. Conversion of olefins into ketones with mercuric acetate and palladium chloride. *Chem. Commun.* 818–819 (1971).
17. Tamura, Y., Ochiai, H., Nakamura, T. & Yoshida, Z. Arylation and vinylation of 2-carboethoxyethylzinc iodide and 3-carboethoxypropylzinc iodide catalyzed by palladium. *Tetrahedr. Lett.* **27**, 955–958 (1986).
18. Inanaga, J., Hirata, K., Sacki, H., Katsuki, T. & Yamaguchi, M. A rapid esterification by mixed anhydride and its application to large-ring lactonization. *Bull. Chem. Soc. Jpn* **52**, 1989–1993 (1979).
19. Arhart, R. J. & Martin, J. C. Sulfuranes. V. The chemistry of sulfur (IV) compounds. Dialkoxydiarylsulfuranes. *J. Am. Chem. Soc.* **94**, 4997–5003 (1972).
20. Hoveyda, A. H., Evans, D. A. & Fu, G. C. Substrate-directable chemical reactions. *Chem. Rev.* **93**, 1307–1370 (1993).
21. Nicolaou, K. C., McGarry, D. G. & Sommers, P. K. New synthetic strategies for the construction of medium-size cyclic ethers. Sterecontrolled synthesis of the BCD ring framework of brevetoxin A. *J. Am. Chem. Soc.* **112**, 3696–3697 (1990).
22. Corey, E. J. & Beames, D. J. Mixed cuprate reagents of type $\text{R}_1\text{R}_2\text{CuLi}$ which allow selective group transfer. *J. Am. Chem. Soc.* **94**, 7210–7211 (1972).
23. Ley, S. V., Norman, J., Griffith, W. P. & Marsden, S. P. Tetrapropylammonium persulfate, $\text{Pr}_4\text{N}^+\text{RuO}_4^-$, TPAP: a catalytic oxidant for organic synthesis. *Synthesis* 639–666 (1994).
24. Nicolaou, K. C. et al. New synthetic technology for the construction of 9-membered ring cyclic ethers. Construction of the EFGH ring skeleton of brevetoxin A. *J. Am. Chem. Soc.* **119**, 8105–8106 (1997).
25. Mahoney, W. S., Brestensky, D. M. & Stryker, J. M. Selective hydride-mediated conjugate reduction of α,β -unsaturated carbonyl compounds using $[(\text{Ph}_3\text{P})\text{CuH}]_6$. *J. Am. Chem. Soc.* **119**, 8105–8106 (1997).
26. Feixas, J., Capdevila, A. & Guerrero, A. Utilization of neutral alumina as a mild reagent for the selective cleavage of primary and secondary silyl ethers. *Tetrahedron* **50**, 8539–8550 (1994).
27. Nicolaou, K. C. et al. Novel strategies for the construction of complex polycyclic ether frameworks. Sterecontrolled synthesis of the FGHJ ring system of brevetoxin A. *Angew. Chem., Int. Edn. Engl.* **30**, 299–303 (1991).
28. Parikh, J. R. & Doering, W. von E. Sulfur trioxide in the oxidation of alcohols by dimethyl sulfoxide. *J. Am. Chem. Soc.* **89**, 5505–5507 (1967).
29. Clayden, J. & Warren, S. Sterecontrol in organic synthesis using the diphenylphosphoryl group. *Angew. Chem., Int. Edn. Engl.* **35**, 241–270 (1996).
30. Zagorski, M. G., Nakanishi, K., Qin, G.-W. & Lee, M. S. Assignment of ^{13}C NMR peaks of brevetoxin A: application of two-dimensional Hartmann-Hahn spectroscopy. *J. Org. Chem.* **53**, 4156–4157 (1988).
31. Lee, M. S., Nakanishi, K. & Zagorski, M. G. Full proton assignment of BTX-A, $\text{C}_{49}\text{H}_{70}\text{O}_{13}$. *New J. Chem.* **11**, 753–756 (1987).
32. Klindgren, B. O. & Nilsson, T. Preparation of carboxylic acids from aldehydes (including hydroxylated benzaldehydes) by oxidation with chlorite. *Acta Chem. Scand.* **27**, 888–890 (1973).
33. Dess, D. B. & Martin, J. C. Readily accessible 12-1-5 oxidant for the conversion of primary and secondary alcohols to aldehydes and ketones. *J. Org. Chem.* **48**, 4155–4156 (1983).
34. Schreiber, J., Maag, H., Hashimoto, N. & Eschenmoser, A. Dimethyl(methylene)ammonium iodide. *Angew. Chem., Int. Edn. Engl.* **10**, 330–331 (1971).
35. Nicolaou, K. C. et al. Total synthesis of brevetoxin B. 1. First generation strategies and new approaches to oxepane systems. *J. Am. Chem. Soc.* **117**, 10227–10238 (1995).
36. Nicolaou, K. C. et al. Total synthesis of brevetoxin B. 2. Second generation strategies and construction of the dioxepane region [DEFG]. *J. Am. Chem. Soc.* **117**, 10239–10251 (1995).
37. Nicolaou, K. C. et al. Total synthesis of brevetoxin B. 3. Final strategy and completion. *J. Am. Chem. Soc.* **117**, 10252–10263 (1995).
38. Nicolaou, K. C. The total synthesis of brevetoxin B: a twelve year odyssey in organic synthesis. *Angew. Chem., Int. Edn. Engl.* **35**, 589–607 (1996).
39. Homaka, Y. Recent methods for detection of seafood toxins: recent immunological method for ciguatera and related polyethers. *Food Addit. Contam.* **10**, 71–82 (1993).
40. Anderson, D. M. Turning back the harmful red tide. *Nature* **388**, 513–514 (1997).
41. Barker, R. *And the Waters Turned to Blood* (Simon & Schuster, New York, 1997).
42. Anderson, D. M. Red tides. *Sci. Am.* **271**, 62–68 (1994).
43. Hall, S. & Strickartz, G. (eds) (American Chemical Society, Washington DC, 1990).
44. Okaichi, T., Anderson, D. M. & Nemoto, T. (eds) *Red Tides. Biology, Environmental Science, and Toxicology* (Elsevier, New York, 1989).
45. Anderson, D. M. & White, A. W. Marine biotoxins at the top of the food chain. *Oceanus* **35**, 55–61 (1992).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) and as paper copy from Mary Sheehan at the London editorial office of Nature.

Acknowledgements. We are indebted to those of our collaborators whose early contributions on this project made its success possible; their names will appear in the full account of this work. We thank K. Nakanishi and Y. Shimizu for samples of natural brevetoxin A, and D. H. Huang, G. Siuzdak and R. Chadha for the NMR, mass spectroscopic and X-ray crystallographic assistance, respectively. This work was supported by the National Institutes of Health USA (GM), The Skaggs Institute for Chemical Biology, Novartis, Merck, Hoffmann-La Roche, Schering Plough, DuPont Merck, the American Chemical Society (graduate fellowship sponsored by Aldrich to J.L.G.) and the Foundation for the Promotion of Scientific Investigations (postdoctoral fellowship to P.G.). Z.Y. and G.S. contributed equally to this project.

Correspondence and requests for materials should be addressed to K.C.N. at The Scripps Research Institute (e-mail: kcn@scripps.edu).

value of 30 at 528 kyr, the lowest peak for a sapropel, the next low value being 37 for sapropel 11 at 409 kyr. Sapropel A thus appears to be the signature of a massive odd monsoon (MOM).

Various data sets can be used to infer the climatic conditions during the deposition of sapropel A. The $\delta^{18}\text{O}$ value of the surface-dwelling foraminifer *Globigerinoides ruber* mainly records the global ice-volume effect on the $\delta^{18}\text{O}$ values of sea water and also local changes in sea surface temperature and salinity¹⁴. Light, depleted values signal high temperature and/or low salinity. In core KC01b, the sea surface temperatures, estimated by alkenones unsaturation ratio, are very low ($14.2 \pm 1.4^\circ\text{C}$ mean value), close to that of glacial sapropel 6 ($14.0 \pm 1.5^\circ\text{C}$), whereas that of interglacial sapropel 5 is $19.8 \pm 1.2^\circ\text{C}$. A cold temperature ($14.3 \pm 1.4^\circ\text{C}$ mean value) is also found in the same sapropel in ODP site 964 ($36^\circ 15' \text{N}$, $17^\circ 44' \text{E}$, 3,672 m water depth) near core KC 01b. These results are corroborated by the weak relative abundance of the warm species *G. ruber*¹⁵. Consequently, the $\delta^{18}\text{O}$ depletion during sapropel A is related to the huge salinity lowering. The climate during the deposition of sapropel A is documented by the simplified pollen record (Fig. 3a). Figure 3b allows comparison with those of sapropel 5, interglacial, and 8, glacial. The pollen abundance for deciduous trees, mainly the deciduous oak (*Quercus*), is high and signals wet conditions and no summer drought¹⁶ in the mid-uplands (about 700 m to 1,200 m above sea level) of the source area, southern Greece and Italy. The pollen abundance of the sagebrush (*Artemisia*) indicates semi-arid conditions in the lowlands. Coolness may have been extant at all elevations. The precipitation increase with elevation is unusually steep.

Thus, sapropel A reflects a large freshwater discharge under dry and cool conditions, at least at sea level. Sea-level aridity also characterizes the younger glacial sapropels 6 (176 kyr) and 8 (220 kyr)^{4,10}, but at those dates, higher elevations were dry as well. In contrast, from the tropical insolation viewpoint, the monsoon index of sapropel A is much lower than those (respectively 51 and 63) during the two glacial sapropels. The occurrence of MOM at that date is paradoxical, particularly as most sapropels predictable by higher insolation are missing. Moreover, the same monsoon index at 151 kyr did not generate a sapropel.

Unexpectedly, this Mediterranean MOM turns out to have a

match in another ocean, which confirms its age and interpretation. From the equatorial Indian Ocean, deep-sea core MD 900963 has yielded a high-resolution $\delta^{18}\text{O}$ record that displays an exceptional depletion event labelled Y at 25.61 m (ref. 5). The isotopic values are generally lower and the excursions of larger amplitude than in other ocean sites. They reflect variations of sea surface salinity at the junction of the Bay of Bengal and the Arabian Sea, controlled by the changing evaporation-minus-precipitation balance of the Indian southwestern monsoon rainfall¹⁷, discharged by the Ganges, Brahmaputra, Irrawady and Salween rivers¹⁸. The timescale, established by tuning to the ice-volume model^{6,7} as for the Mediterranean record, has been facilitated by the strong expression of the precession in this monsoon domain. Figure 2c, d shows the detailed setting for event Y in that core and in nearby core MD 900961. Event Y is a $\delta^{18}\text{O}$ depletion peak (-2.55‰), dated 525 kyr. It is followed by a heavy peak (-0.55‰) labelled X at 24.91 m, dated 510 kyr, between the two interglacial isotopic stages 13 and 15. The Y-X succession has no correspondence in other Indian Ocean records, which have a lesser resolution^{19,20}. Taking into account their good chrono-stratigraphical agreement, sapropel A and the Y event in cores KC01b and MD900963 seem to be the signature of the same climatic event.

Thus, the thick Mediterranean sapropel A was deposited at 528–525 kyr under low salinity of the cold surface water, while a massive fluvial discharge spilled over the Indian Ocean, despite the fact that heavy monsoonal precipitation in India and Africa would not be predicted by the insolation conditions. The Earth passed at perihelion on the June solstice just before 528,000 yr BP, while obliquity had an average value and eccentricity was almost at a minimum⁷. None of the orbital parameters had a remarkable value. Nevertheless, the double manifestation of MOM suggests that a common cause at a large regional scale has been at work. At this point, it is important to assess the scope of its source area. Was MOM linked to unusual circumstances in one or in two oceans?

Mediterranean sapropels seem to have been triggered by the stratification of the water column, mainly due to the swelled discharge of the Nile River flood³. The sea surface can be stably stratified by the plume of a suddenly enlarged influx of fresh water, while winds are weak and mixing of the sea surface is minimal. The Nile discharge increases tenfold between June and August²¹ until

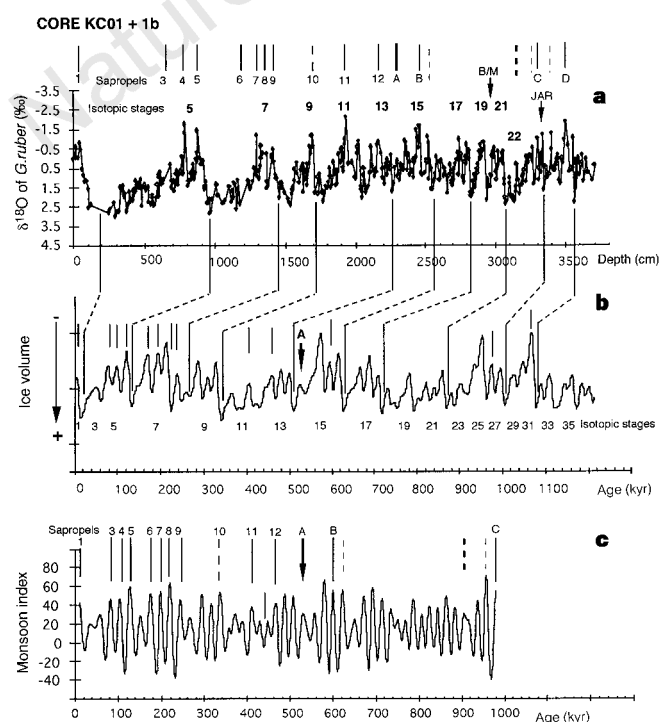


Figure 1 The large-scale framework of sapropel A. **a**, $\delta^{18}\text{O}$ record versus depth of *G. ruber* (white) for Ionian Sea core KC01b ($36^\circ 15.25' \text{N}$, $17^\circ 44.34' \text{E}$; 3,643 m water depth, 3,704 cm long) at 5-to-10 cm intervals, with sapropels (vertical lines) and ghosts (dotted lines) identified by increased ratio of Ba/Al and low intensity of the anhysteretic remanent magnetization^{8,13,28}. For the first 250 cm, the data are from nearby core KET 82 22 ($37^\circ 56' \text{N}$, $16^\circ 53' \text{E}$; 3,643 m water depth) because the first two metres of core KC01b are disturbed. Magnetic events are recorded: B/M, Brunhes/Matuyama reversal; JAR, top of Jaramillo event. **b**, Ice-volume model for the past 1.2 Myr (ref. 6). The age model of core KC01b has been established in two steps, first by visually correlating the ten main cold stadials in the two data sets. Second, after a Blackman–Tukey spectral analysis of this first rough chronology, the two data sets have been filtered at the same frequency (0.043 kyr^{-1} , period 23 kyr), and tuned up until a good fit was obtained between the two filtered signals, with $R = 0.87$. **c**, Monsoon index as calculated according to ref. 7 (insolation $23^\circ \text{N} + \delta$ insolation $23^\circ \text{N} - 0^\circ$, caloric half year) (ref. 4) for the past 1 Myr. The astronomical solution of ref. 7 for the ice-volume model was used for tuning the three isotopic records and for the monsoon index.

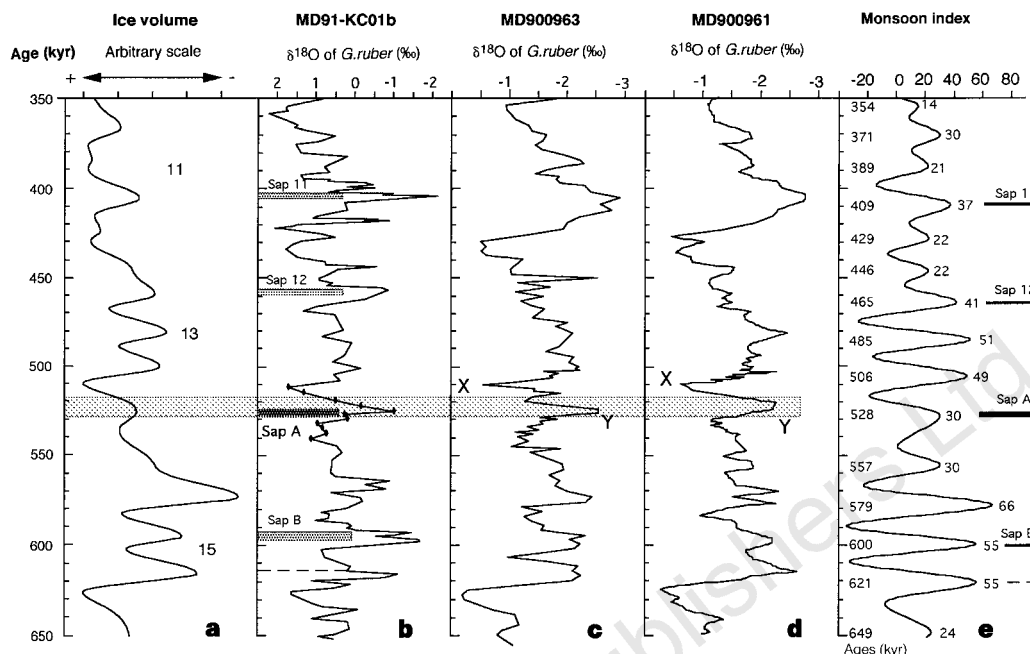


Figure 2 The setting for sapropel A in detailed records for the interval from 350 to 650 kyr, isotopic stages 11 to 15. The grey area underlines the same isotopic event. **a**, Ice-volume model⁶. **b**, Timescale of the $\delta^{18}\text{O}$ record of *G. ruber* (white) for core KC01b, obtained by tuning to the ice-volume model. Within sapropel A, there were three $\delta^{18}\text{O}$ measurements: 0.18‰ at 2,300.5 cm, 0.26‰ at 2,296 cm and -1.02‰ at 2,294.5 cm. Individual measurements are signalled by diamonds; black within the sapropel, white below and above. **c**, Timescale of the $\delta^{18}\text{O}$ record of *G. ruber* (white) at 10-cm intervals for giant piston core MD 900963 (5° 03.30' N, 73° 52.60' E; 2,446 m water depth) in the equatorial Indian Ocean, east of the Maldives platform, at the junction of the Bay of Bengal and the Arabian Sea⁵. Event X at 510 kyr is

synchronous with the cold stadial between stages 13 and 15, and event Y at 525 kyr with sapropel A. **d**, Timescale of the $\delta^{18}\text{O}$ record (10-cm intervals) for nearby core MD 900961 (5° 03.71' N, 73° 52.57' E; 2,445 m water depth), also displaying the X and Y events. Both timescales in **c** and **d** were obtained by tuning to the ice-volume model^{6,7}. **e**, The monsoon index. The sapropel ages calculated by using the ice-volume model (**b**) generally lag behind those estimated by using the monsoon index as discussed previously²⁷ (**e**). This is because the ice-volume model introduces phase lags of ~5 kyr between the insolation and the $\delta^{18}\text{O}$ in the precession bands⁶, whereas monsoon index chronology assumes an immediate response of sapropels to insolation forcing³.

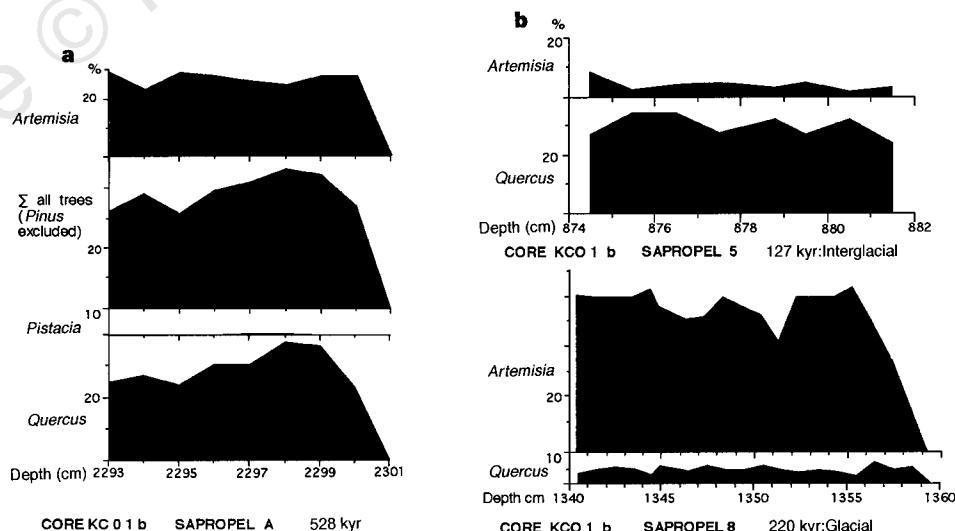


Figure 3 Mediterranean climatic conditions. **a**, The pollen record of sapropel A at 1-cm sampling resolution. Pollen grain count was 450 to 600 per level, with mean *Pinus* abundance being 25–30%. In the sapropel, the high pollen concentration (50,000–70,000 grains g^{-1}) reveals a good preservation of all pollen types. Percentages of selected types are shown, calculated on a pollen sum from which *Pinus* is excluded. *Quercus*, the deciduous oak, is a high pollen producer in the deciduous forest belt at ~700–1,200 m elevation¹⁶ during interglacials (precipitation was at least 650 mm yr^{-1} , there was no summer drought and winter was neither prolonged nor very harsh). The sum of all trees excludes *Pinus* and includes Mediterranean trees (*Olea* and *Pistacia*) at low-middle

elevations, and temperate deciduous trees above. *Artemisia*, the sage brush growing in semidesert conditions at low elevation, signals that precipitation was below 200 mm yr^{-1} . **b**, Comparison with the simplified pollen records of interglacial sapropel 5 and glacial sapropel 8 shows the intermediate character of sapropel A: the percentage for *Quercus* was similar or higher than in sapropel 5; for *Artemisia*, it was half that in sapropel 8. This distribution in sapropel A is interpreted as being mainly controlled by the available moisture, which is low near the coast and high at mid-elevations. Temperature can be cool at all elevations. The deciduous forest at 528 kyr was extant under lower temperature than during sapropel 5, but the available moisture was above a minimum of 650 mm at both dates.

October, a rise never achieved by temperate rivers. In contrast, the discharge into the Mediterranean of the European rivers mainly fed by winter and spring precipitation and snow melt is subject to strong seasonal wind mixing. Moreover, at least in the Early Holocene, two new climatic conditions in Mediterranean Europe and south-west Asia have contributed to the inhibition of Mediterranean deep-water formation: no-frost winters, inferred from the expansion of the pistachio tree, and moderately increased summer rainfall, inferred from that of the deciduous forest²².

The Blue Nile generates the annual flood. It drains the orographic monsoonal rainfall over the northern Ethiopian highlands. The moist, shallow equatorial westerlies which in summer cover northern tropical Africa²³ originate from the pressure gradient between the high over the subtropical south Atlantic and the northernmost position over the Sahara (20° N) of the trough of low pressure which traces the seasonal course of the Sun. The monthly distribution of $\delta^{18}\text{O}$ in rain in Khartoum and Addis Ababa shows very depleted values in summer which reveal the continental effect from a distant, Atlantic oceanic source of moisture^{24,25}. For the rest of the year, the easterlies bring little but isotopically heavier moisture from the Indian ocean. The formation of sapropels in the eastern Mediterranean is the distant legacy of the winter evaporation from the subtropical South Atlantic. In contrast, the moisture that precipitates as monsoon rainfall over the Indian subcontinent is exported from the subtropical South Indian Ocean. The cause of MOM must have had an extensive scope, acting over two oceans and two continents.

'Ghost' sapropels^{8,28} help evaluate MOM in a wider perspective. They all correspond to higher values of the monsoon index than sapropel A, except one, at 3,160 cm core depth. Dated at 905 kyr (monsoon index timescale), its monsoon index is 29. Its stratigraphic level in relationship to the precession and monsoon index cycles is similar to that of sapropel A: both occur in the second cycle after the highest monsoon index, which is in phase with the peak of each 400-kyr eccentricity cycle, respectively dated at 579 kyr and 959 kyr. At these dates, however, neither sapropel nor ghost is observed, while the previous cycle, with a lower monsoon index, is respectively marked by sapropels B at 600 kyr and C at 980 kyr. This observation inspires speculation as to the possibility of a large-scale nonlinearity with respect to the 400-kyr eccentricity cycle.

Over Africa and southern Asia, intensification of the precipitation of the oceanic moisture requires a steeper ocean-to-continent pressure gradient²⁶, enhanced by stronger heating over the summer continent. Precisely at 528 kyr, however, neither the insolation over the two continents, nor the intertropical insolation gradient, the same as at present, seem to have been strong enough to enhance the monsoon adequately according to the linear pattern that acted over the last 250 kyr. We suggest that the cause of MOM might lie with other processes, such as a non linear response to insolation. □

Received 23 May 1997; accepted 15 January 1998.

- Kullenberg, B. On the salinity of the water contained in marine sediments. *Meddel. Fr. Oceanog. Instit. i Göteborg* **21**, 1–38 (1952).
- Olausson, E. Studies of deep sea cores. Sediment cores from the Mediterranean Sea and the Red Sea. *Rep. Swed. Deep Sea Exped. 1947–48* **8**, 337–391 (1961).
- Rossignol-Strick, M., Nesteroff, W., Olive, P. & Vergnaud-Grazzini, C. After the deluge: Mediterranean stagnation and sapropel formation. *Nature* **295**, 105–110 (1982).
- Rossignol-Strick, M. African monsoon, an immediate climate response to orbital insolation. *Nature* **304**, 46–49 (1983).
- Bassinot, F. et al. The astronomical theory of climate and the age of the Brunhes-Matuyama magnetic reversal. *Earth Planet. Sci. Lett.* **126**, 91–108 (1994).
- Imbrie, J. & Imbrie, J. Z. Modeling the climatic response to orbital variations. *Science* **207**, 943–953 (1980).
- Berger, A. Long-term variations of caloric insolation resulting from the Earth's orbital elements. *Quat. Res.* **9**, 139–167 (1978).
- Langereis, C. G., Dekkers, M. J., de Lange, G. J., Paterne, M. & van Sandvoort, P. J. M. Magnetotratigraphy and astronomical calibration of the last 1.1 Myr from an Eastern Mediterranean piston core and dating of short events in the Brunhes. *Geophys. J. Int.* **129**, 75–94 (1997).
- Ryan, W. B. F. in *The Mediterranean Sea: A Natural Sedimentation Laboratory* (ed. Stanley, D. J.) 149–169 (Dowden, Hutchinson and Ross, Stroudsburg, PA, 1972).
- Vergnaud-Grazzini, C., Ryan, W. B. F. & Cita, M. B. Stable isotopic fractionation, climate change and episodic stagnation in the Eastern Mediterranean during the Late Quaternary. *Mar. Micropal.* **2**, 353–370 (1977).

- Higgs, N. C., Thomson, J., Wilson, T. R. S. & Croudace, I. W. Modification and complete removal of Eastern Mediterranean sapropels by postdepositional oxidation. *Geology* **22**, 423–426 (1994).
- Thomson, J. et al. Redistribution and geochemical behaviour of redox-sensitive elements around S1, the most recent Eastern Mediterranean sapropel. *Geochim. Cosmochim. Acta* **59**, 3487–3501 (1995).
- Van Os, B. J. H., Middleburg, J. J. & de Lange, G. J. Possible diagenetic mobilization of barium in sapropelic sediment from the Eastern Mediterranean. *Mar. Geol.* **100**, 125–136 (1991).
- Shackleton, N. J. in *Les Méthodes Quantitatives d'Étude des Variations du Climat au Cours du Pléistocène 203–209* (Colloque Intern. CNRS No 219, Paris, 1974).
- Sanvoisin, R., D'Onofrio, S., Lucchi, R., Violanti, D. & Castradori, D. 1 Ma paleoclimatic record from the Eastern Mediterranean-MARFLUX project: first results of a micropaleontological and sedimentological investigation of a long piston core from the Calabrian Ridge. *Il Quaternario* **6**, 169–188 (1993).
- Beug, H. J. Changes of climate and vegetation belts in the mountains of Mediterranean Europe during the Holocene. *Bull. Geol. (Warszawa)* **19**, 101–110 (1975).
- Rostek, F. et al. Reconstructing sea surface temperature and salinity using $\delta^{18}\text{O}$ and alkenone records. *Nature* **364**, 319–321 (1993).
- Duplessy, J. C. Glacial to interglacial contrasts in the northern Indian Ocean. *Nature* **295**, 494–498 (1982).
- Peterson, L. C. & Prell, W. L. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations, Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 251–269 (AGU, Washington DC, 1985).
- Divakar Naidu, P., Malmgren, B. A. & Bornmalm, L. Quaternary history of calcium carbonate fluctuations in the western equatorial Indian Ocean (Somali basin). *Palaeogeogr. Palaeoclimat. Palaeoecol.* **103**, 21–30 (1993).
- Shahin, M. *Hydrology of the Nile Basin* (Elsevier, Amsterdam, 1985).
- Rossignol-Strick, M. Sea-land correlation of pollen records in the Eastern Mediterranean for the Glacial-Interglacial transition: biostratigraphy versus radiometric time-scale. *Quat. Sci. Rev.* **14**, 893–915 (1995).
- Newell, R., Kidson, J. W., Vincent, D. G. & Boer, G. J. in *The General Circulation of the Tropical Atmosphere and Interactions with Extratropical Latitudes* Vol. 1, 87 and 105 (MIT Press, Cambridge, MA, 1972).
- Gat, J. R. & Gonfiantini, R. in *Statistical Treatment of Data on Environmental Isotopes in Precipitation* 352, 356 (Tech. Rep. ser. 331, IAEA, Vienna, 1992).
- Rozanski, W., Araguas-Araguas, L. & Gonfiantini, R. in *Climate Change in Continental Isotopic Records* (eds van der K., Lohman, K. C., McKenzie, J. & Savin, S.) 1–36 (Geophys. monogr. 78, AGU, Washington, 1993).
- Kutzbach, J. E. Monsoon climate of the Early Holocene: climate experiment with the Earth's orbital parameters for 9000 years ago. *Science* **214**, 59–61 (1981).
- Hilgen, F. J., Lourens, L. J., Berger, A. & Loutre, M. F. Evaluation of the astronomically calibrated time scale for the late Pliocene and earliest Pleistocene. *Paleoceanography* **8**, 549–565 (1993).
- van Santvoort, P. J. M., de Lange, G. J., Langereis, C. G., Dekkers, M. J. & Paterne, M. Geochemical and paleomagnetic evidence for the occurrence of "missing" sapropels in eastern Mediterranean sediments. *Paleoceanography* **12**, 773–786 (1997).
- Berger, A. & Loutre, M. F. Insolation values for the climate of the last 10 million years. *Quat. Sci. Rev.* **10**, 297–317 (1991).
- Hilgen, F. J. Astronomical calibration of Gauss to Matuyama sapropels in the Mediterranean and implications for the geomagnetic polarity time-scale. *Earth Planet. Sci. Lett.* **104**, 226–244 (1991).

Acknowledgements. The multidisciplinary investigations of core KC01b have been supported by the European Community Marine Science and Technology (MAST I and II) contracts Marflux and Paleoflux. M.R.-S. was supported by CNRS; M.P. and E.B. were supported by CEA and CNRS. Hand-picking of foraminifers was by courtesy of C. Corselli, Milan University; isotope facilities were provided by L. Labeyrie at CFR, Gif-sur-Yvette.

Correspondence and requests for materials should be addressed to M.R.-S. (e-mail: mrstrick@ccr.jussieu.fr).

Abyssal hills formed by stretching oceanic lithosphere

W. Roger Buck* & Alexei B. Poliakov†

* Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA

† Laboratoire de Géophysique et Tectonique, CNRS UMR 5573, Université Montpellier II, Montpellier Cedex 05, France

Tectonic plates are formed and move apart at mid-ocean ridges. Some portion of this plate-separation process can occur by stretching of the crust, resulting in a complex pattern of extensional faults. Abyssal hills, the most ubiquitous topographic features on Earth¹, are thought to be a product of this faulting^{2,3}. Here we report the results of a self-consistent numerical model of lithospheric formation and stretching that includes spontaneous fault creation. In this model, an axial valley develops where the fault activity is most concentrated. The 'frozen' fault-generated topography, rafted out of the axial valley, is visually and statistically similar to observed abyssal hills formed at many slower-spreading ridges. Faults appear to be replaced by new faults because their offset changes the local stress field. We accordingly need no temporal variation in magmatism, as required by some

previous models⁴⁻⁶, to control the spacing or offset of faults. Our model results suggest instead that the irregularity of abyssal hill relief may result from a self-organized critical stress state at spreading centres.

The large-scale 'axial valley' relief of many slow- to intermediate-spreading ridges (for example, as shown in Fig. 1) is thought to result from lithospheric stretching²⁻⁹. Previous numerical models of mid-ocean-ridge stretching do not resolve localized shear zones or model faults. Although these models successfully reproduce steady-state axial valley shape¹⁰ they cannot simulate abyssal-hill formation.

Recent models for abyssal-hill formation^{6,11} have considered elastic deformation around predefined faults. In these models the underlying cause of the spacing and the amount of offset on these faults has to be assumed. Our model treats large-scale brittle and viscous deformation which can result in large-scale structures, such as axial valleys, but also allows for spontaneous development of small-scale faults when a failure criterion is met. The main problem is to explain what controls the spacing and offset of normal faults at ridges.

Several authors argue that magmatism is a critical factor for abyssal-hill formation, because magmatic crustal accretion by dyking can occur at stress levels that are lower than that required for normal faulting^{4,5,6,12}. In their view, surface faulting can only develop during periods of reduced magma supply. Small abyssal hills imply frequent, short amagmatic periods whereas large abyssal hills imply infrequent, long amagmatic periods. Irregularity in faulting might reflect irregularity in magmatism at a ridge. Unfortunately, we have neither observations of magma supply variations nor testable models of processes controlling these assumed variations.

In previous work¹³ we used numerical models to show that continuous magmatic accretion of oceanic crust need not suppress

the surface expression of normal faults developing in response to mantle lithospheric stretching. In those calculations we only considered the initiation and small offset of such faults. In the present calculations we allow unlimited separation of model lithospheric plates.

We attempt to produce the simplest possible model that results in both an axial valley and abyssal hills. In this end-member model, processes of magmatic accretion of the crust are not treated. A two-dimensional region, with a laterally variable thermal structure, is stretched with boundary conditions shown in Fig. 2. The temperature field shown in this figure is an approximate representation of the way steady-state temperatures vary spatially in thermal models of spreading centres^{14,15}. Temperatures do not change during a calculation: material moves through a fixed temperature field.

The shallow, cold part of the model region behaves as a brittle material, approximated with Coulomb elasto-plastic rheology. Warmer regions deform by thermally activated creep, approximated as linear visco-elastic flow. We assume that the viscosity depends only on temperature (Fig. 2).

Faults only form in brittle materials which become weaker when they break. In our model, brittle deformation is localized in 'fault zones' owing to our assumption of a strain-dependent and strain-rate-dependent reduction in shear strength, or cohesion. The reduction of cohesion during deformation is represented as $S = S_0 \exp(-\epsilon^*/\epsilon_r)$, where S_0 is the cohesive strength, ϵ^* is the local effective plastic strain, and ϵ_r is an assumed constant. Numerical difficulties related to artificial diffusion of strain led us to introduce a kind of healing of inactive faults. Thus, the rate of change of the effective plastic strain is given by $d\epsilon^*/dt = d\epsilon/dt - \epsilon^*/\tau$, where τ is a time constant, and $d\epsilon/dt$ is the total strain rate.

We use an explicit finite-element method similar to the FLAC (fast lagrangian analysis of continua) technique developed by Cundall¹⁶. It has been used to simulate localized deformation

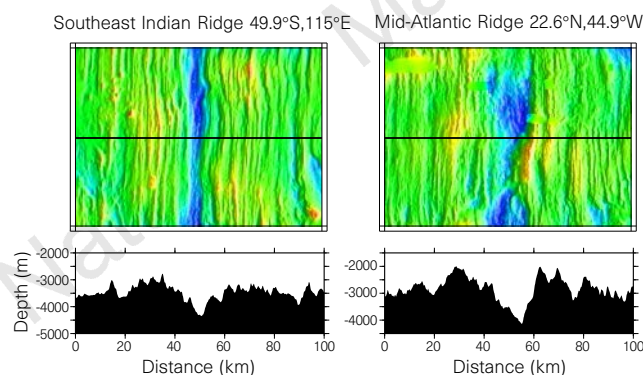


Figure 1 Ridge-parallel abyssal-hill topography. These shaded relief bathymetric maps (top row) show segments of the intermediate-spreading-rate Southeast Indian ridge (data from Cochran *et al.*²⁸) and the slow-spreading-rate Mid-Atlantic Ridge (data from Gente *et al.*²⁹). Shallower depths are in warmer colours. The scaling between depth and colour is different for the two maps. The maps are oblique Mercator projections, centred on the locations given at the top of the map and rotated so that the axial valley and abyssal hill relief trend in the local north-south direction. Solid lines in the bathymetric maps show the location of profiles shown in the bottom row. The abyssal hills show along-strike continuity for tens of kilometres (that is, greater than the lithospheric thickness) suggesting that two-dimensional models may be relevant to the processes forming these features.

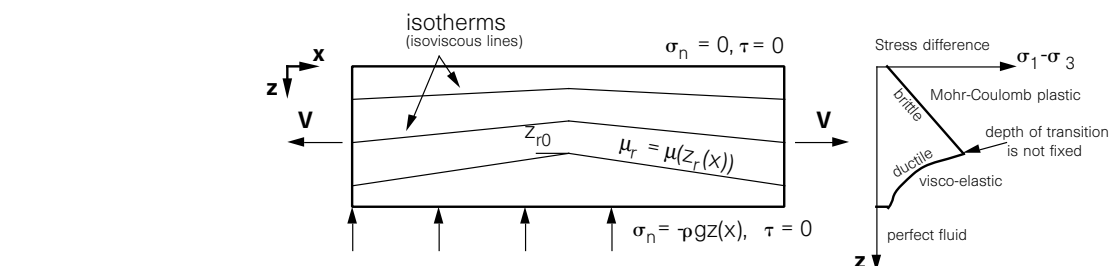


Figure 2 Details of model used. Boundary conditions and isotherms are shown on the left; a representative lithospheric yield stress profile is shown on the right. The temperature field is assumed to be steady-state with the depth to a given isotherm increasing linearly with distance, x , from the spreading centre. We approximate temperature-dependent rheology by taking the viscosity: $\mu(x, z) = \mu_r \exp(-(z - z_r)/L_e)$, where μ_r is the viscosity fixed for a given geotherm

at a reference depth $z_r(x) = z_r(0) + \alpha x$, where α is a constant and z_e is the distance over which the viscosity changes by a factor e . We note that the transition between brittle and ductile behaviour is not prescribed. V is boundary velocity, σ_n is normal stress, σ_1 and σ_3 are the maximum and minimum principal stresses.

(approximating faults) in elasto-plastic materials applied to a variety of problems^{17–19}.

This method is lagrangian (that is, the numerical grid follows the deformation), so the simulation of very large deformation involves re-meshing to overcome the problem of degradation of numerical precision when elements are distorted. We have tested different criteria to trigger re-meshing and find that the essential results presented here do not depend on the amount of strain between re-meshing.

The model results in development of an evolving system of faults. The pattern of instantaneous strain rate for one of our calculations is shown in Fig. 3. Areas of localized high strain rate in the shallow portion of the model axis are shear zones that we refer to as model faults. For both of the model times shown in the figure, the amount of boundary separation was large compared to the brittle-layer thickness at the axis. At different times in the calculation, the pattern of faulting is different.

The evolution of model topography through time (Fig. 4) shows that the first feature formed is a valley, generated by one and then a pair of faults dipping in towards the axis. The next faults dipped outwards, away from the axis. Eventually, faults formed at a variety of distances from the axis in what appears to be a chaotic fashion.

An axial valley is a consistent feature of this model case, but the shape of the valley changes with time. Smaller-scale and more-

random topographic relief is formed within and at the edge of this axial valley. This abyssal-hill-like relief is rafted away from the axis without further significant deformation.

Statistical properties of the model topography compare well with similar properties estimated for sea-floor topography. To avoid the topography of the model axial valley, we analysed only the region more than 30 km from the model axis in the last profile shown in Fig. 4 (labelled 208 km). Long-wavelength topography was removed by subtracting a second-order polynomial fit through the profile on each side of the axis. The 'roughness' of the abyssal hills was estimated in terms of the square root of the mean of the squared topography. We found values of 176 m and 210 m for roughness of the left and right side profiles, respectively. An autocorrelation analysis was used to find a characteristic width of the topographic relief of 7.5 km for the left side, and 10 km for the right side, profiles. The values of both the roughness and characteristic width fall in the middle of the range estimated for abyssal hills close to the Mid-Atlantic Ridge²⁰.

Varying several of the model parameters significantly affected the predicted topographic relief. For zero cohesion (that is, $S_0 = 0$) we do not get even small abyssal-hill-like relief. We still get an axial valley, but its shape does not change (detectably) with time.

The axial-valley relief correlates with the geometry and strength of the brittle layer in the model. This brittle geometry is not a

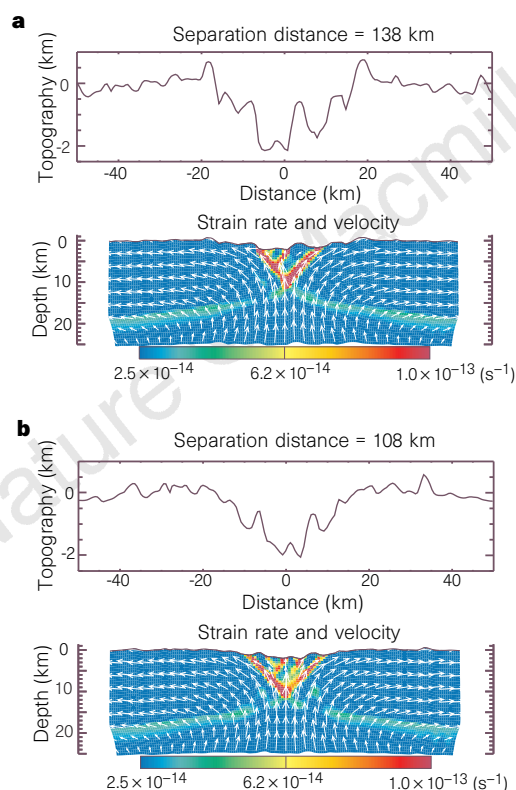


Figure 3 Cross-section of strain rate (in colour with scale at base of image), velocity (as white arrows) and topography (with vertical exaggeration of 10) for two amounts of horizontal offset in a representative calculation. Elasticity is controlled by the Lamé moduli, both of which equal 3×10^{10} Pa. We assume incompressible plastic flow with yielding defined by a friction coefficient of 0.6, $S_0 = 20$ MPa, $\epsilon_r = 0.3$, $\tau = 1.0 \times 10^{12}$ s. The viscosity field is described by $\mu_r = 10^{22}$ Pa s, $z_r(x) = 10$ km, $\alpha = 0.1$ and $z_e = 1$ km. The half-spreading velocity is 5×10^{10} m s⁻¹, (~ 1.5 cm yr⁻¹). We show only the central, 100-km-wide, part of the calculation domain which had an initial width of 200 km and depth of 50 km. The brittle deformation at the axis extends down to ~ 12 km at the axis. This is the depth where the viscous stresses and the brittle stresses are the same. The pattern of high strain rate, or faulting, is very different for the two times shown.

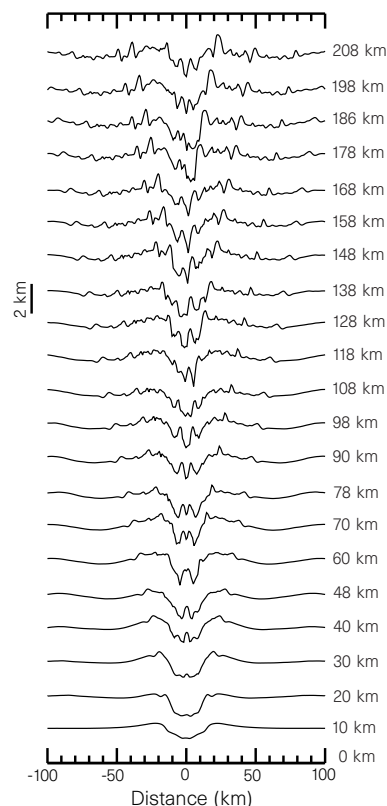


Figure 4 Sequence of model topographic profiles for the same case illustrated in Fig. 3. The pattern of topography, in terms of whether it is axial-valley-like or abyssal-hill-like, varies in an irregular way after a start-up period when topography was symmetric and regular. Here the vertical exaggeration is 9.

directly input parameter, but it depends on the imposed temperature field in a fairly linear way. Doubling the layer thickness approximately doubles the width and depth of the valley. Increasing the rate of thickening of the isotherms with distance from the axis increases the depth of the axial valley.

Though we described the change of cohesion in shear zones with three parameters, some simple relations result. The maximum abyssal-hill relief appears to scale with the amount of cohesion drop in a given shear zone. As noted, when the cohesion drop is zero then no abyssal hills are seen. When the cohesion drop is large, compared to the average stress difference needed for frictional slip, then the abyssal hills are larger relative to the axial valley. For extreme cases, the axial valley is not clearly discernible in the model relief. This may reflect the same type of material behaviour used in a recent analogue model of lithospheric stretching in which discrete faults develop, but an axial valley is not a dominant topographic feature²¹.

Our model topography may be understood at a simple mechanical level by the idea that normal fault offset is self-limiting²². The offset of a normal fault produces topographic relief and bending, both of which require work, and changes the stress field around the fault²³. Increasing the offset makes it harder to continue slip on a given fault and facilitates the breaking of a new fault. Based on this kind of reasoning, we expected to see a regular pattern of faulting and topography.

As well as fault offset being limited, we also saw irregular behaviour in cases with finite cohesion. It is this feature that makes model topography look like observed abyssal-hill relief. We do not think that this variability is a numerical artefact as it is found only for certain physical parameters. Also, thorough studies of shear localization in a brittle material similar to that considered here, but with simpler boundary conditions, found irregular behaviour for a wide range of numerical parameters²⁴. These results may reflect the fact that the axial valley is in a critical state and relieves the stress by simultaneous slipping on a system of faults in a way that can be described as self-organized criticality²⁵.

Our main conclusion is that abyssal-hill-like topography may result from continuous stretching of a brittle layer. Other processes such as crustal accretion almost certainly do affect faulting and development of topography at ridges, but the underlying cause of abyssal hills at ridges with axial valleys may be brittle stretching of cold mantle lithosphere. As we do not include magmatic crustal accretion and spreading-rate-dependent thermal structures, we cannot yet assess the controls on the amplitude of abyssal hills for the full range of spreading conditions. We expect that ratio of crustal thickness to brittle-mantle thickness at the axis may be an important factor controlling abyssal-hill size.

Abyssal hills formed at fast-spreading ridges are much smaller in amplitude and appear to form in a very different way from those at slow-spreading ridges²⁶. These features are probably not explained by the kind of lithospheric stretching modelled here. The fact that fast-spreading ridges lack an axial valley may reflect the absence of brittle mantle lithospheric stretching¹⁰. The small abyssal hills associated with fast-spreading ridges may result from bending and breaking of brittle upper-crustal lithosphere in response to the vertical loads of topography and extrusion of magma²⁷. □

Received 12 May 1997; accepted 6 January 1998.

- Menard, H. W. *Marine Geology of the Pacific* (McGraw-Hill, New York, 1964).
- MacDonald, K. C. & Luyendyk, B. Deep-tow studies of the structure of the Mid-Atlantic Ridge near 37°N (FAMOUS). *Geol. Soc. Am. Bull.* **88**, 621–636 (1977).
- Shaw, P. R. Ridge segmentation, faulting and crustal thickness in the Atlantic. *Nature* **358**, 490–493 (1992).
- Karson, J. A. *et al.* Along-axis variations in seafloor spreading in the MARK area. *Nature* **328**, 681–685 (1987).
- Malinverno, A. & Pockalny, R. A. Abyssal hill topography as an indicator of episodicity in crustal accretion. *Earth Planet. Sci. Lett.* **99**, 154–169 (1990).
- Thatcher, W. & Hill, D. P. A simple model for fault-generated morphology of slow spreading ridges. *J. Geophys. Res.* **100**, 561–570 (1995).
- Tapponnier, P. & Francheteau, J. Necking of the lithosphere and the mechanics of slowly accreting plate boundaries. *J. Geophys. Res.* **83**, 3955–3970 (1978).

- Phipps Morgan, J., Parmentier, E. M. & Lin, J. Mechanisms for the origin of mid-ocean ridge topography: Implications for the thermal and mechanical structure of accreting plate boundaries. *J. Geophys. Res.* **92**, 12823–12836 (1987).
- Lin, J. & Parmentier, E. M. A finite amplitude necking model of rifting in brittle lithosphere. *J. Geophys. Res.* **95**, 4909–4923 (1990).
- Chen, Y. & Morgan, W. J. A nonlinear rheology model for mid-ocean ridge axis topography. *J. Geophys. Res.* **95**, 17583–17604 (1990).
- Shaw, W. J. & Lin, J. Model of ocean ridge lithospheric deformation: Dependence on crustal thickness, spreading rate, and segmentation. *J. Geophys. Res.* **101**, 17977–17993 (1996).
- Kappel, E. S. & Ryan, W. B. F. Volcanic episodicity and non-steady state rift valley along northeast Pacific spreading centers: Evidence from Sea MARC I. *J. Geophys. Res.* **91**, 13925–13940 (1986).
- Polakov, A. N. B. & Buck, W. R. in *Faulting and Magmatism at Mid-Ocean Ridges* (AGU Monograph, Am. Geophys. Union, in the press).
- Morton, J. L. & Sleep, N. H. A mid-ocean ridge thermal model: Constraints on the volume of axial hydrothermal flux. *J. Geophys. Res.* **90**, 11345–11353 (1985).
- Lin, J. & Parmentier, E. M. Mechanisms of lithosphere extension at mid-ocean ridges. *Geophys. J. Int.* **96**, 1–22 (1989).
- Cundall, P. A. Numerical experiments on localization in frictional materials. *Ingenieur-Archiv.* **59**, 148–159 (1989).
- Hobbs, B. E., Muhlaus, H. B. & Ord, A. *Instability, Softening and Localization of Deformation* 143–165 (Spec. Publ. 54, Geol. Soc., London, 1990).
- Polakov, A. N. B. & Herrmann, H. J. Self-organized criticality in plasticity shear bands. *Geophys. Res. Lett.* **21**, 2143–2146 (1994).
- Hassani, R. & Chery, J. Control of extensional tectonics by crustal rheology: numerical experiments. *Geology* **24**, 1095–1098 (1996).
- Goff, J. A. Near-ridge abyssal hill morphology. *J. Geophys. Res.* **98**, 21713–21737 (1991).
- Shemenda, A. L. & Grocholsky, A. L. Physical modeling of slow seafloor spreading. *J. Geophys. Res.* **99**, 9137–9153 (1994).
- Forsyth, D. W. Finite extension and low-angle normal faulting. *Geology* **20**, 27–30 (1992).
- Buck, W. R. Effect of Lithospheric thickness on the formation of high- and low-angle normal faults. *Geology* **21**, 933–936 (1992).
- Polakov, A. N. B., Hermann, H. G., Podladchikov, Y. & Roux, S. Fractal plastic shear bands. *Fractals* **2**, 567–581 (1994).
- Bak, P., Tang, C. & Wiesenfeld, K. Self-organized criticality. *Phys. Rev. A* **38**, 364–374 (1988).
- MacDonald, K. C. *et al.* Volcanic growth faults and the origin of Pacific abyssal hills. *Nature* **380**, 125–129 (1996).
- Buck, W. R. Bending thin lithosphere causes localized “snapping” and not distributed “crunching”: Implications for abyssal hill formation. *Geophys. Res. Lett.* **24**, 2531–2534 (1997).
- Cochran, J. R., Semper, J.-C. & the SEIR Scientific Team The Southeast Indian Ridge between 88°E and 118°E: Gravity anomalies and crustal accretion at intermediate spreading rates. *J. Geophys. Res.* **102**, 15463–15488 (1997).
- Gente, P. *et al.* Characteristics and evolution of the segmentation of the Mid-Atlantic Ridge between 20°N and 24°N during the last 10 million years. *Earth Planet. Sci. Lett.* **129**, 55–72 (1995).

Acknowledgements. We thank J. Lin and J. Goff for reviews. This work was supported by NSF and CNRS.

Correspondence and requests for materials should be addressed to W.R.B. (e-mail address: buck@lamont.ldeo.columbia.edu).

The skull of a relative of the stem-group bird *Mononykus*

Luis M. Chiappe*, Mark A. Norell† & James M. Clark‡

* Department of Ornithology, and † Department of Vertebrate Paleontology, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024, USA

‡ Department of Biological Sciences, George Washington University, Washington DC, 20052, USA

In joint expeditions, researchers from the American Museum of Natural History and the Mongolian Academy of Sciences have recovered over 20 alvarezsaurid (Theropoda: Aves) specimens in the Late Cretaceous beds of Mongolia's Gobi Desert¹. Here we describe a new taxon that is closely related to *Mononykus*^{2,3}. This new taxon is represented by two exquisitely preserved skulls—the first known for Alvarezsauridae—details of which support the theory that the group is related to birds^{4,5}. This theory was first put forward on the basis of primarily postcranial evidence^{2,3}, including the presence of avian characteristics such as the absence of a contact between the jugal and postorbital, and between the quadratojugal and squamosal, articulations. Other earlier evidence that suggested that the alvarezsaurids were birds included the presence of a movable joint between the quadratojugal and quadrate, separate squamosal and braincase articulations of the quadrate, confluence between the caudal tympanic recess and columellar recess, a triradiate palatine, an unusually large foramen

magnum, and the loss of a coronoid bone. The configuration of the temporal region of the skull and its articulation with the rostrum indicate the capability for prokinetic movement in which flexing occurs at the junction of the upper jaw and neurocranium, and support the idea that prokinesis preceded other types of avian intracranial kinesis.

Aves (Avialae *sensu* Gauthier 1986) L., 1758

Metornithes Perle *et al.* 1993

Alvarezsauridae Bonaparte 1991

Mononykinae tax. nov.

Shuvuuia deserti gen. et. sp. nov.

Etymology. *Shuvuuia* (from Mongolian 'shuvuu'): bird; *deserti*, in reference to the semi-arid depositional environment of the Djadokhta Formation¹.

Holotype. MGI 100/975, Mongolian Geological Institute, Ulaanbaatar.

Referred material. MGI N 100/99, MGI 100/1001. These specimens, including the holotype, were previously referred to *Mononykus*⁴.

Locality and horizon. Ukhaa Tolgod (MGI 100/975, MGI 100/977, MGI 100/1001) and Tugrugen Shireh (MGI 100/99), South Gobi Aimak, Mongolia. Tentatively Djadokhta Formation (Late Cretaceous/Campanian?)^{1,6}.

Diagnosis. A mononykine distinguished from *Mononykus olecranus* by its less compressed cervical centra with large pneumatic foramina, humeral deltopectoral crest that is continuous with its head, pubis of subcircular section, femoral and tibiotarsal shafts bowed latero-medially, and medial margin of astragalus' ascending process less excavated. *Shuvuuia deserti* shows less co-ossification of the proximal tarsals to the tibia and among metacarpals in specimens of comparable size to the holotype of *Mononykus olecranus*, suggesting a lower rate of bone co-ossification. The new species differs from *Parvicursor remotus* because it lacks a ventral keel in the most rostrally located synsacral vertebra and because it has less co-ossification between proximal tarsals and the tibia. *Shuvuuia deserti* differs from all alvarezsaurids because it has a sharp ridge on the medial margin of the distal tibiotarsus, a character that we have interpreted to be an autapomorphy (a novel character of a terminal taxon that evolved from an existing character). The skull of *Shuvuuia deserti* shows several characteristics that we have interpreted as autapomorphies, including the articulation between quadrate and postorbital, the elongated basiptyergoid processes, the numerous teeth, and the hypertrophied prefrontal/ectethmoid (see below for a discussion of the problematic interpretation of this bone). These characters, however, may prove to diagnose a more inclusive taxon after the discovery of additional alvarezsaurid cranial material.

The delicate skull is elongate with large orbits and terminal tear-shaped nares (Figs. 1–3). The maxilla forms most of the snout's lateral surface as in other early birds⁷. The caudal half forms a slender bar bordering the ventral side of the antorbital fossa. The dentigerous margin of the premaxilla is not preserved but the maxilla bears numerous tiny, unserrated teeth, identical to those of *Mononykus*². Teeth are absent behind the rostral third of the antorbital fossa. The teeth lay in a continuous groove on the dentary.

The lacrimal has an expanded rostral process and a tiny caudal extension resembling the L-shaped lacrimal of *Archaeopteryx* and other non-avian theropods⁸. The orbit is walled rostrally by a large ossification, widely exposed on the skull roof. This ossification has been difficult to identify. It could be interpreted as a large prefrontal, but both elements meet ventrally to the frontals—a condition unknown in dinosaurs⁹. In this respect, this element is more similar to an avian ectethmoid, but no ectethmoid is known to be exposed dorsally¹⁰.

The postorbital fails to reach the rod-like jugal, so the orbit and infratemporal fossa are continuous; this condition is known only

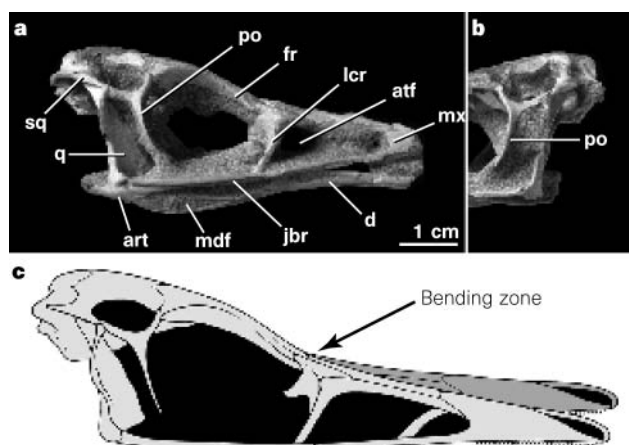


Figure 1 Skull and jaw of *Shuvuuia deserti* (MGI 100/1001). **a**, Right lateral and **b**, left lateral (only the postorbital region is shown with the left jaw removed) views. **c**, Reconstruction of the skull of *Shuvuuia deserti* displaying its kinematic capacity for elevating the snout independently of the braincase. Placement of flexion areas or bending zones at the base of the snout, both dorsally and ventrally, suggests that the snout moved as a single unit, as in all prokinetic birds. Note that *Shuvuuia*'s prokinetic movement differs from the modern type of prokinesis in that the dorsal bending zone lies between the frontals and the nasal-preorbital bones, and not between the frontals and the nasal-premaxillaries. The shaded area indicates the area the snout can move to. Abbreviations: art, articular; atf, antorbital fossa; d, dentary; fr, frontal; jbr, jugal bar; lcr, lacrimal; mdf, mandibular fossa; mx, maxilla; po, postorbital; q, quadrate; sq, squamosal.

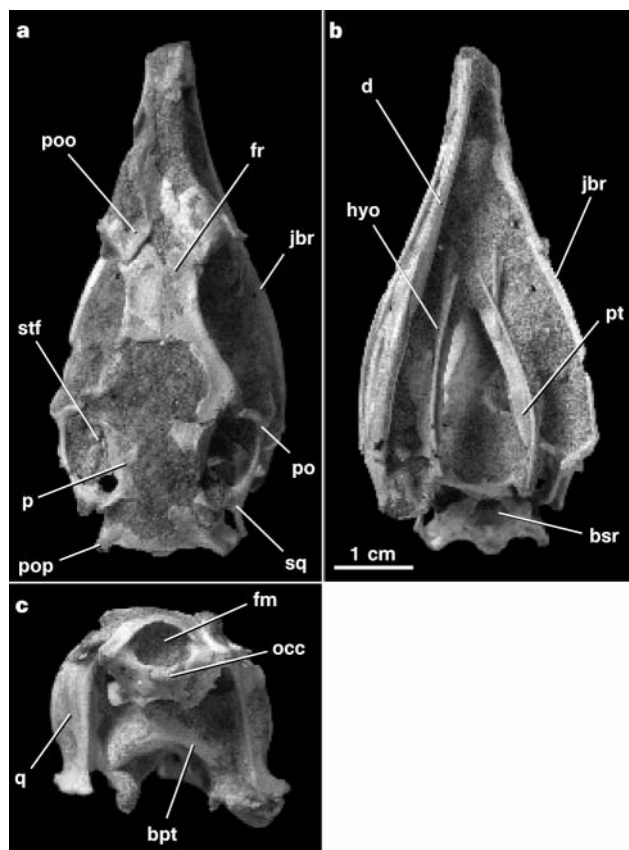


Figure 2 Skull and right jaw of *Shuvuuia deserti* (MGI 100/1001) in dorsal (**a**), ventral (**b**), and caudal (**c**) views. Abbreviations: bpt, basiptyergoid process; bsr, basisphenoidal recess; fm, foramen magnum; hyo, hyoid; occ, occipital condyle; p, parietal; poo, preorbital ossification (prefrontal/ectethmoid); pop, paroccipital process; pt, pterygoid; stf, supratemporal fenestra; other abbreviations are as in Fig. 1.

for metornithine birds among archosaurs^{11,12}, although it may be present in *Archaeopteryx*⁴. The jugal and quadratojugal are fused into a jugal bar that forms a tiny fork on its caudal end. *Shuvuuia deserti* is also different from non-avian archosaurs, in which the quadratojugal has a tall dorsal process that is sutured to the quadrate^{11,12}. In *Archaeopteryx*¹³ and an enantiornithine hatchling¹⁴ this dorsal process is shorter and is not sutured to the quadrate. The triradiate squamosal is not incorporated into the braincase of

Shuvuuia, *Archaeopteryx*, and other early avians, which retain a complete, diapsid supratemporal fenestra^{13,14}.

As in other birds, the paroccipital processes are short and are not perforated by the caudal tympanic recess, which opens inside the columellar recess⁴. The foramen magnum is proportionally larger than in any non-avian dinosaur⁹ and the caudally directed occipital condyle is very small. The basiptyergoid processes are unusually long and almost vertically oriented.

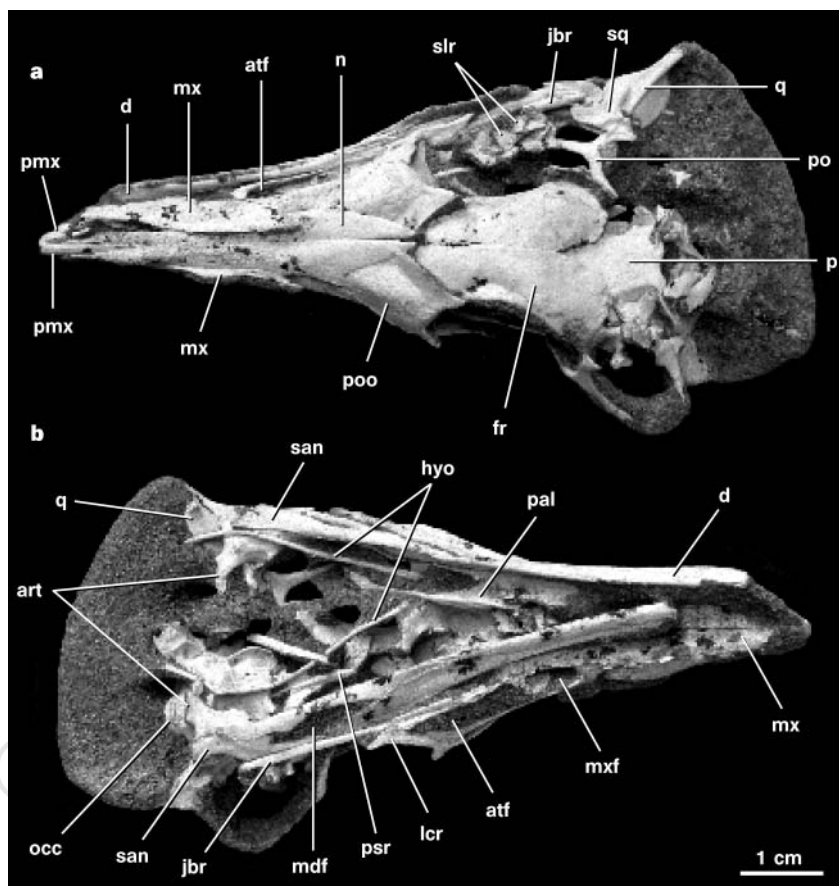


Figure 3 Skull and jaws of *Shuvuuia deserti* (MGI 100/977) in dorsal (a) and ventral (b) views. Abbreviations: mxl, maxillary fenestra; n, nasal; pal, palatine;

pmx, premaxilla; psr, parasphenoidal rostrum; san, surangular; slr, sclerotic ring; other abbreviations as in Figs 1 and 2.

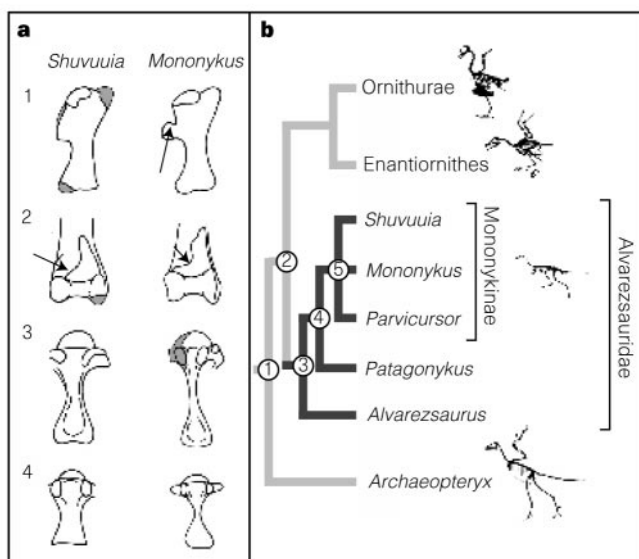


Figure 4 Taxonomy of *Shuvuuia deserti* and of Alvarezsauridae. **a**, Some morphological differences between *Shuvuuia deserti* and *Mononykus olecranus* (figures not in scale). **1**, *Shuvuuia* has a cranio-dorsally expanded deltopectoral crest that is continuous with the humeral head whereas *Mononykus* exhibits a pillar-like deltopectoral crest (arrow). **2**, *Shuvuuia* possesses a sharp ridge (longer arrow) on the medial margin of the distal end of the tibia that is absent in *Mononykus* and lacks the deep medial notch (shorter arrow) at the base of the ascending process of *Mononykus*. **3**, **4**, The centra of *Shuvuuia*'s cervico-dorsal (**3**) and anterior dorsals (**4**) are much less compressed than those in *Mononykus*. For autapomorphic characters and further differences see Diagnosis in text. **b**, Phylogenetic position of the Alvarezsauridae and inter-relationships among its component taxa. This is a consensus cladogram derived from the three most parsimonious trees with different topologies for the mononykine taxa under the implicit enumeration command ('finds all shortest trees') of computer program 'Hennig 86' (length = 115 steps; consistency index = 0.83; RI : 0.82 for both the consensus and the fundamental cladograms). *Mononykinae* is phylogenetically defined as the common ancestor of *Mononykus*, *Shuvuuia*, and *Parvicursor*, plus all their descendants. Diagnoses of nodes 1-5 are provided in the Supplementary information.

The quadrate is tall and slender, resembling the quadrates of *Archaeopteryx* and the enantiornithine *Gobipteryx* in that it is nearly one-quarter the length of the skull¹³. This bone also resembles that of *Archaeopteryx* in the strong rostro-caudal compression of its latero-distal corner¹³. Proximally, the quadrate has two well differentiated heads. The lateral head articulates with a squamosal-postorbital facet, whereas the medial head is dorso-caudally directed toward the braincase. The palatine is long and slender, lacking the jugal process and tetraordinate aspect of non-avian theropod palatines¹³.

Study of the skull of *Shuvuuia* provides further evidence for the avian affinities of the Alvarezsauridae and emphasizes the highly specialized nature of this bizarre lineage: *Shuvuuia* displays a number of unusual cranial (as well as postcranial) characters and some characters that are restricted to Aves among dinosaurs. The configuration of the jugal and suspensorium of *Shuvuuia* (Fig. 1) suggests a capacity for intracranial kinesis^{12,15–17} (for example, the elevation and depression of the rostrum). Without ventral squamosal and dorsal quadratojugal processes, the streptostylic quadrate would have been free to swing antero-posteriorly^{12,15–17}. The lack of a connection between the jugal and postorbital would have freed the jugal to act as a strut between the quadrate and rostrum. Forces directed longitudinally from the quadrate would rotate the rostrum around a transverse axis at a flexion area just anterior to the orbit. A thinning of the jugal (bending zone) just caudal to its lacrimal contact and the loose connection between the frontals and the preorbital bones (nasals and prefrontals/ectethmoids) indicate that the snout may have moved as a unit like in prokinetic birds^{12,15–17} (Fig. 1c). This interpretation is supported by the absence of a continuous naso-orbital septum. Prokinesis is usually regarded as the primitive avian type of kinesis derived from either akinetic or mesokinetic archosaurian skulls^{12,15–17} and has been seen in several early birds^{18–20}. The design of the skull of *Shuvuuia* suggests that some motion, probably prokinetic, was possible, supporting the theory that prokinesis was a primitive type of kinesis.

A cladistic analysis based on 90 characters (six of which are multistate) places the Alvarezsauridae as the sister taxon to all avians except for *Archaeopteryx* (Fig. 4; see Supplementary information for character list, data matrix, and node diagnoses). Cranial characters of *Shuvuuia* that are shared only with birds among dinosaurs include the absence of a postorbital-jugal contact, the movable joint (which is not sutured) between the quadratojugal and quadrate, the separate articulation of the quadrate with the braincase, and the disproportionately large foramen magnum relative to the occipital condyle. The skulls of *Shuvuuia*, *Archaeopteryx*, and other avians share characters that are absent in velociraptorine theropods, the outgroup selected for this cladistic analysis. These characters include the absence of a squamosal-quadratojugal contact and a coronoid in the mandible, and the presence of a triradiate palatine, a caudal tympanic recess confluent with the columellar recess, and unserrated tooth crowns.

It has been claimed that alvarezsaurids are either specialized ornithomimosaurs^{21,22} or a different group of non-maniraptoran theropods²³, although in neither case has a cladistic analysis been published. Our phylogenetic studies and independent cladistic analyses indicate that the few similarities between alvarezsaurids and non-maniraptoran coelurosaurs are the result of convergent evolution^{4,5,24–26}. For example, the ornithomimosaur *Pelecanimimus* bears numerous tiny teeth, with unserrated crowns, that are restricted to the anterior part of the maxilla²⁷, and ornithomimosaurs (as well as oviraptorosaurs and therizinosaurids among non-avian maniraptorans) lack a coronoid bone²⁸. Furthermore, if the pre-orbital ossification of *Shuvuuia* is identified as a prefrontal, it is larger than that of most maniraptoran dinosaurs and more comparable in size to that of ornithomimosaurs²⁹. However, placing alvarezsaurids outside Maniraptora would require homoplasy in an extensive number of maniraptoran, avian, and metornithine synapomorphies^{4,5,24–26}. □

Received 10 September; accepted 26 November 1997.

1. Dashzeveg, D. et al. Extraordinary preservation in a new vertebrate assemblage from the Late Cretaceous of Mongolia. *Nature* **374**, 446–449 (1995).
2. Perle, A., Norell, M. A., Chiappe, L. M. & Clark, J. M. Flightless bird from the Cretaceous of Mongolia. *Nature* **362**, 623–626 (1993).
3. Perle, A. et al. Skeletal morphology of *Mononykus olecranus* (Theropoda: Avialae) from the Late Cretaceous of Mongolia. *Am. Mus. Novit.* **3105**, 1–29 (1994).
4. Chiappe, L. M., Norell, M. A. & Clark, J. M. Phylogenetic position of *Mononykus* from the Upper Cretaceous of the Gobi Desert. *Mem. Queens. Mus.* **39**, 557–582 (1996).
5. Chiappe, L. M., Norell, M. A. & Clark, J. M. *Mononykus* and birds: methods and evidence. *Auk* **114**, 300–302 (1997).
6. Jerzykiewicz, T. & Russell, D. A. Late Mesozoic stratigraphy and vertebrates from the Gobi Basin. *Cretac. Res.* **12**, 345–377 (1991).
7. Chiappe, L. M. Late Cretaceous birds of southern South America: anatomy and systematics of *Enantiornithes* and *Patagopteryx deferrariisi*. *Münchener Geowiss. Ab. A* **30**, 203–244 (1996).
8. Currie, P. J. New information on the anatomy and relationships of *Dromaeosaurus albertensis* (Dinosauria, Theropoda). *J. Vert. Paleontol.* **15**, 576–591 (1995).
9. Weishampel, D. B., Dodson, P. & Osmólska, H. *The Dinosauria* (Univ. of California Press, Berkeley, 1990).
10. Cracraft, J. The lacrimal-ectethmoid bone complex in birds: a single character analysis. *Am. Midl. Nat.* **80**, 316–359 (1968).
11. Romer, A. S. *Osteology of the Reptiles* (Univ. Chicago Press, 1956).
12. Zusi, R. L. in *The Skull Vol. 2* (eds Hanken, J. & Hall, B. L.) 391–437 (Univ. Chicago Press, 1993).
13. Elzanowski, A. & Wellnhofer, P. Cranial morphology of *Archaeopteryx*: evidence from the seventh skeleton. *J. Vert. Paleontol.* **16**, 81–94 (1996).
14. Sanz, J. L. et al. A nesting bird from the Early Cretaceous of Spain: implications for avian skull and neck evolution. *Science* **276**, 1543–1546 (1997).
15. Zusi, R. L. A functional and evolutionary analysis of rhynchokinesis in birds. *Smithson. Contr. Zool.* **395**, 1–40 (1985).
16. Simonetta, A. M. On the mechanical implications of the avian skull and their bearing on the evolution and classification of birds. *Q. Rev. Biol.* **35**, 206–220 (1960).
17. Bock, W. J. Kinetics of the avian skull. *J. Morphol.* **114**, 1–42 (1964).
18. Bühler, P. in *The Beginnings of Birds* (eds Hecht, M. K., Ostrom, J. H., Viohl, G. & Wellnhofer, P.) 135–140 (Proc. Int. Archaeopteryx Conf., Eichstätt, 1985).
19. Hou, L., Martin, L. D., Zhou, Z. & Feduccia, A. Early adaptive radiation of birds: evidence from fossils from northeastern China. *Science* **274**, 1164–1167 (1996).
20. Bühler, P., Martin, L. D. & Witmer, L. M. Cranial kinesis in the Late Cretaceous birds *Hesperornis* and *Parahesperornis*. *Auk* **105**, 111–122 (1988).
21. Martin, L. D. & Rinaldi, C. How to tell a bird from a dinosaur. *Maps Digest* **17**, 190–196 (1994).
22. Feduccia, A. *The Origin and Evolution of Birds* (Yale Univ. Press, New Haven, 1996).
23. Sereno, P. The origin and evolution of dinosaurs. *Ann. Rev. Earth Planet. Sci.* **25**, 435–489 (1997).
24. Novas, F. E. Alvarezsauridae, Cretaceous maniraptorans from Patagonia and Mongolia. *Mem. Queens. Mus.* **39**, 675–702 (1996).
25. Novas, F. E. Anatomy of *Patagonykus puertai* (Theropoda, Maniraptora, Alvarezsauridae) from the Late Cretaceous of Patagonia. *J. Vert. Paleontol.* **17**, 137–166 (1997).
26. Forster, C. A., Chiappe, L. M., Krause, D. W. & Sampson, S. D. The first Cretaceous bird from Madagascar. *Nature* **382**, 532–534 (1996).
27. Pérez-Moreno, B. P. et al. A unique multitoothed ornithomimosaur from the Lower Cretaceous of Spain. *Nature* **370**, 363–367 (1994).
28. Clark, J. M., Perle, A. & Norell, M. A. The skull of *Erlcosaurus andrewsi*, a Late Cretaceous “Segnosaur” (Theropoda: Therizinosauridae) from Mongolia. *Am. Mus. Novit.* **3115**, 1–39 (1994).
29. Osmólska, H., Roniewicz, E. & Barsbold, R. A new dinosaur, *Gallimimus bullatus* n. gen. n. sp. (Ornithomimidae) from the Upper Cretaceous of Mongolia. *Palaeontol. Pol.* **27**, 103–143 (1972).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

Acknowledgements. We thank A. Davidson and M. Ellison for the preparation and illustration of the specimens, respectively, and A. Milner, L. Witmer, G. Zweers, and J. Vanden Berge for useful reviews and discussions. The Chapman and Frick Memorial Funds of the AMNH and the National Science Foundation provided support for this research.

Common reference frame for neural coding of translational and rotational optic flow

D. R. W. Wylie, W. F. Bischof & B. J. Frost*

Department of Psychology, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

* Department of Psychology, Queen's University, Kingston, Ontario K7L 3N6, Canada

Self-movement of an organism through the environment is guided jointly by information provided by the vestibular system and by visual pathways that are specialized for detecting 'optic flow'^{1,2}. Motion of any object through space, including the self-motion of organisms, can be described with reference to six degrees of freedom: rotation about three orthogonal axes, and translation along these axes. Here we describe neurons in the pigeon brain that respond best to optic flow resulting from

translation along one of the three orthogonal axes. We show that these translational optic flow neurons, like rotational optic flow neurons³⁻⁵, share a common spatial frame of reference with the semicircular canals of the vestibular system. The three axes to which these neurons respond best are the vertical axis and two horizontal axes orientated at 45° to either side of the body midline.

As the environment contains many stationary objects and surfaces, self-motion induces distinctive patterns of visual motion ('optic flow' or 'flowfields') across the entire retina⁶. Psychophysical research has illustrated the importance of optic flow for the control of posture and locomotion^{7,8} and for the perception of self-motion⁹. Figure 1b, c shows flowfields resulting from both self-translation (Fig. 1c) and self-rotation (Fig. 1b). The shaded areas in Fig. 1c indicate differences in local motion in the translational flowfield. At one 'pole' in the direction of translation, flow radiates outward from the focus of expansion, while converging to the focus of contraction (not shown) behind the bird's head, with laminar at the 'equator'. In this figure, the bird is translating forward, along the z-axis. As the pigeon has laterally placed eyes, forward translation results in backward (that is, nasal to temporal) visual motion throughout much of the visual field of both eyes.

Neurons sensitive to translational and rotational optic flow have been found in the visual neuropile of the blowfly¹⁰ and, in vertebrates, in extrastriate visual cortex^{11,12}, the accessory optic system (AOS) and the vestibulocerebellum^{3-5,13}. In pigeons, the AOS¹⁴ and the vestibulocerebellum¹³ contain neurons that are responsive to either translational or rotational flowfields. Here we show that the translation-sensitive neurons respond best to translational optic flow along one of three axes, namely, the vertical axis (y-axis) or one of two horizontal axes orientated at 45° to the midline.

We recorded the activity of neurons in the nucleus of the basal optic root (nBOR, a component of the AOS) and of Purkinje cells (complex spikes) in the ventral uvula and nodulus of the vestibulocerebellum in anaesthetized pigeons. Cells in these structures have large binocular receptive fields, covering much of the entire visual field of both eyes, and respond best to large moving visual stimuli that approximate to translational flowfields^{13,14}. (Cells in the flocculus

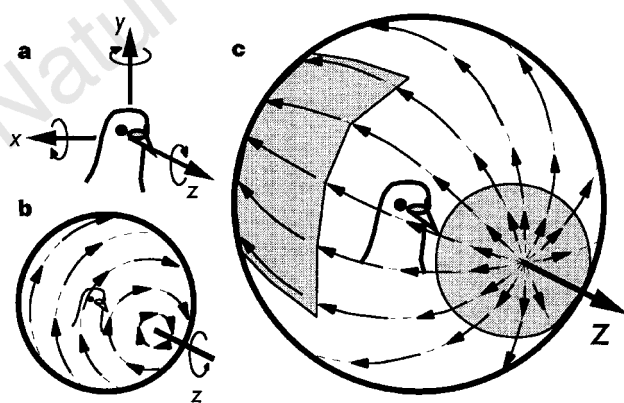


Figure 1 Generic description for motion of an object in three-dimensional space and optic flowfields generated by translation along, and rotation about the z-axis. **a**, The generic description for motion of an object in three-dimensional space using the standard coordinate system. Motion can be described using a reference frame consisting of three orthogonal axes (x, y and z), and six degrees of freedom, three of translation (straight arrows) and three of rotation (curved arrows). This is an egocentric coordinate system, that is, it refers to motion of the bird. **b**, The optic flowfield resulting from a clockwise rotation of the bird about the z-axis. **c**, The optic flowfield resulting from translation (out from the page) along the z-axis. The shaded areas in **c** highlight the differences in local image motion at the 'pole' and 'equator' of the translational flowfield.

of the vestibulocerebellum selectively respond to rotational flowfields^{5,13}.) The binocular neurons in the nBOR represent a small subpopulation: most nBOR neurons are monocular and were not studied here. We projected a translational flowfield onto the walls, ceiling and floor of the room with a specially designed 'translator'. The translator is similar to the 'planetarium projector'^{3,4}

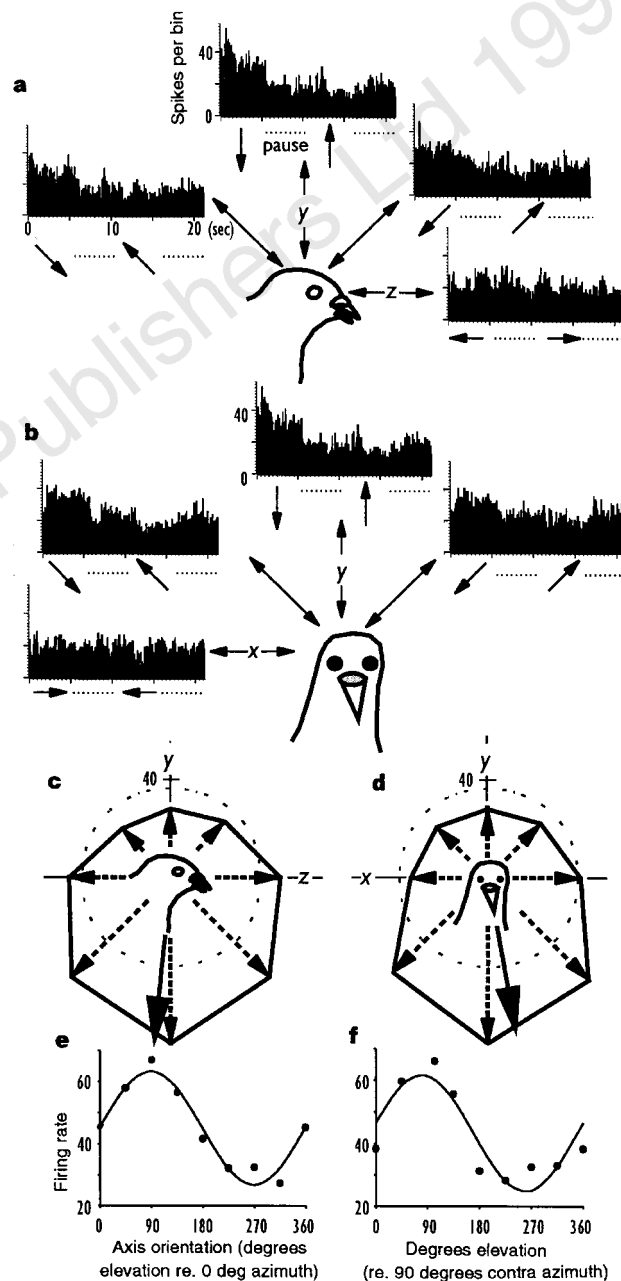


Figure 2 Tuning curves of a binocular nBOR neuron that is maximally responsive to translational optic flow along the vertical axis are shown. **a**, **b**, PSTHs show the responses to translation along axes in the sagittal (**a**) and frontal (**b**) planes. The tuning curves are shown as polar plots for the sagittal and frontal planes in **c** and **d**, respectively, where firing rate (spikes per second) is plotted as a function of the orientation of the direction of translational flow. The orientation of an arrow reflects the orientation of the axis of the translator, and the arrowhead points in the direction in which the bird would be moving to produce such a flowfield. The broken circles represent the spontaneous firing rate, and the solid arrows indicate the best axes as determined from the best cosine fits. **e** and **f** show the best cosine fits to the tuning curves shown in **c** and **d**, respectively.

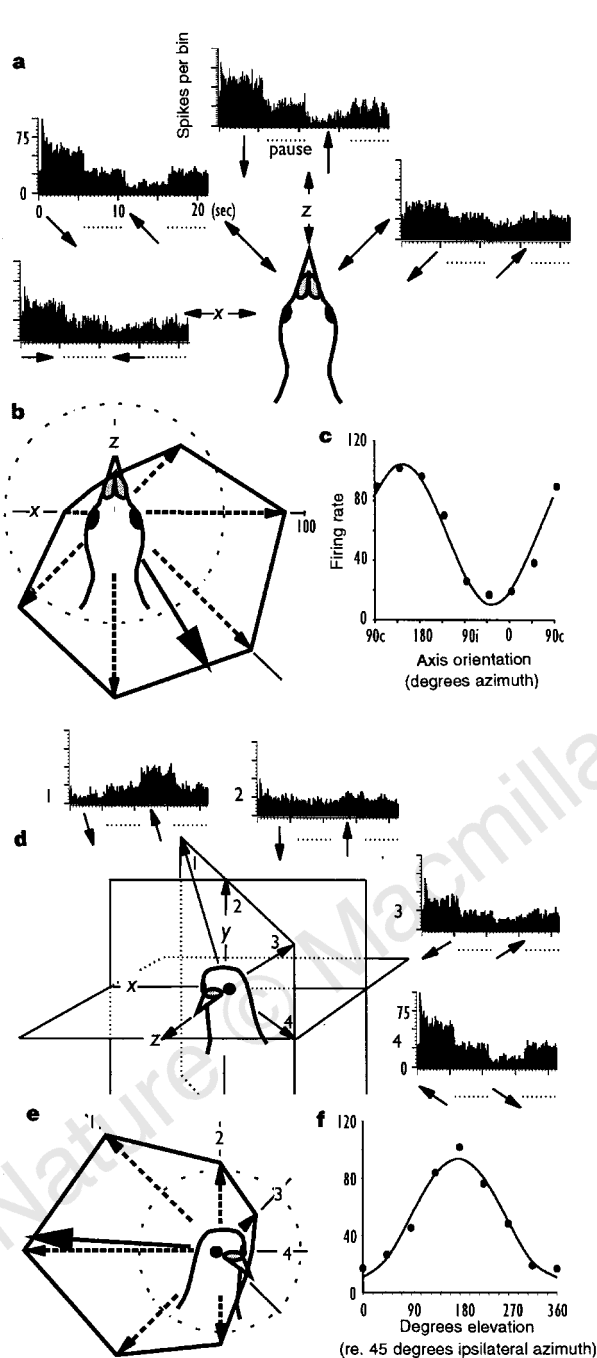


Figure 3 Tuning curves of a binocular nBOR neuron that is maximally responsive to translational optic flow along an horizontal axis orientated $\sim 45^\circ$ to the midline are shown. **a**, PSTHs show the responses of neurons to translation along axes in the horizontal plane. **d**, PSTHs show the responses to translation along axes in a vertical plane that intersects the horizontal plane at 45° ipsilateral azimuth. Polar plots of the tuning curves are shown in **b** and **e**, for the planes depicted in **a** and **d**, respectively. The broken circles represent the spontaneous firing rate, and the solid arrows indicate the best axis as determined from the best cosine fit. **c**, **f**, Show the best cosine fits to the tuning functions shown in **b** and **e**, respectively. 90i, 90° ipsilateral azimuth; 90c, 90° contralateral azimuth.

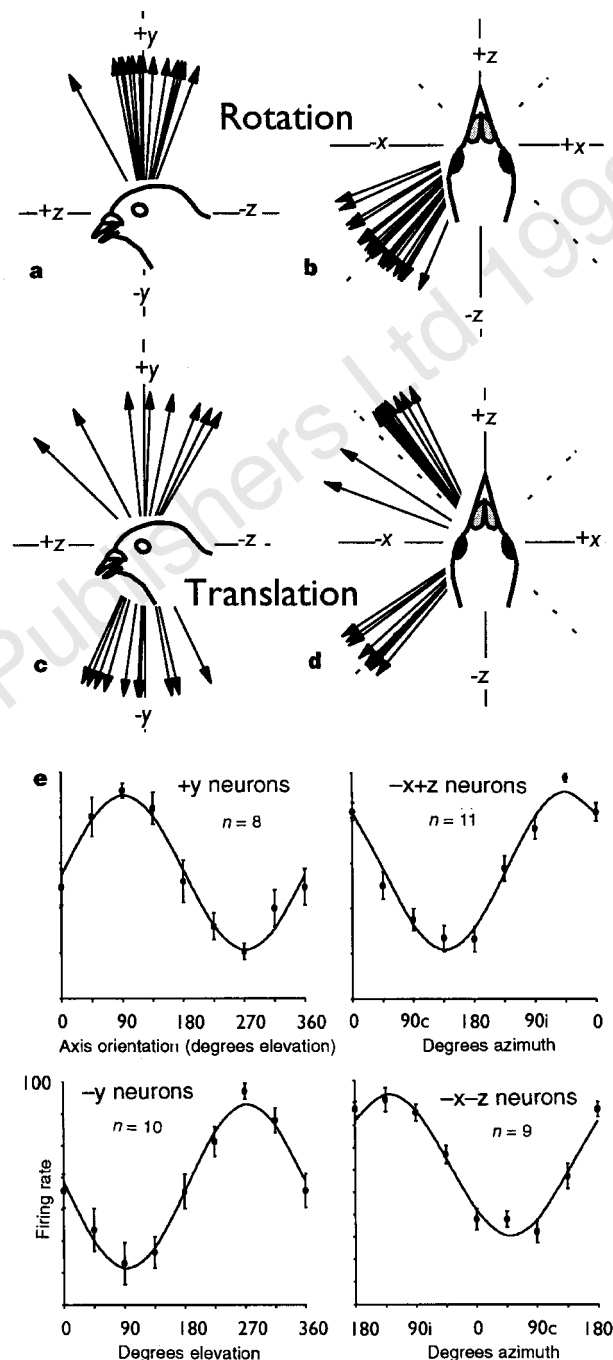


Figure 4 Best axes of translation- and rotation-sensitive neurons in the vestibulocerebellum of pigeons. **a**, **b**, The reference frame of rotation-sensitive neurons in the flocculus (from ref. 5). **c**, **d**, The reference frame of the translation-sensitive neurons in the uvula and nodulus from the present study. In **a**, **b**, each arrow represents the best axis about which a rotational visual flowfield resulted in maximal modulation. The arrows in **c**, **d**, represent the best axes of the translation-sensitive neurons in the uvula and nodulus from the present study. In **a**, **c**, the best axes are projected onto a sagittal plane and were determined from elevation tuning curves in that plane. Likewise, in **b**, **d**, the best axes are projected onto the horizontal plane and were determined from azimuthal tuning curves in that plane. **e**, The best cosine fits to the mean normalized tuning curves¹³ for each of the four groups of translation-sensitive neurons are shown. For these curves, the firing rate in response to translation in each direction is expressed as a percentage of the cell's maximal firing rate, then averaged within groups. Average firing rate ($\pm$ s.e.m.) is plotted as a function of the direction of translation.

in that it presents motion to the entire visual field, but is different in that it simulates translations rather than rotations (see Methods).

We found that although neurons were broadly tuned, they responded best to translational optic flow along a particular axis, but responded minimally to translation along axes orthogonal to this 'best axis'. Along the best axis, motion in one direction resulted in strong excitation, whereas motion in the opposite direction produced inhibition. We recorded data from 67 neurons that responded optimally to translational optic flow (38 neurons in the ventral uvula and nodulus of the vestibulocerebellum and 29 neurons in the nBOR). For simplicity, we describe the data as if all recordings were obtained from the left side of the brain.

Figure 2 shows the responses of an nBOR neuron to translational flowfields along several axes. In response to movement of a large handheld stimulus in the central part of the visual field (see Methods), this neuron was excited in response to upward visual flow in both hemifields but was insensitive to horizontal motion. Figure 2a, b shows peristimulus time histograms (PSTHs) illustrating the responses to translational optic flow in eight directions (45° apart) in both the sagittal (Fig. 2a) and the frontal (Fig. 2b) planes. The data from Fig. 2a, b are shown in Fig. 2c, d, where the average firing rate is plotted (in polar plots) as a function of the direction of translational optic flow. The orientation of an arrow reflects the orientation of the axis of the translator, and the arrowhead points in the direction in which the animal would be moving to produce such a flowfield. That is, the arrowhead points toward the focus of expansion in the flowfield. The solid arrows in the polar plots denote the 'best axis' for the tuning curves in Fig. 2c, d, determined from the best cosine fits shown in Fig. 2e, f. The fits are excellent, indicating that; for the two planes tested, the response magnitude of this neuron to translation in any direction $\mathbf{d}$ was linearly related to the projection of $\mathbf{d}$ onto the neuron's preferred direction. This neuron responded best to $-y$ translation and was strongly inhibited by $+y$ translation. Translational optic flow along the x -axis (Fig. 2b, d) and z -axis (Fig. 2a, c) produced very little modulation.

In contrast, the nBOR neuron shown in Fig. 3 showed little modulation to translation along the y -axis, but responded best to translational optic flow along a horizontal axis orientated at 45° ipsilateral azimuth. This neuron was excited by forward (temporal to nasal) movement of the handheld stimulus in the central region of both hemifields but was insensitive to vertical visual flow. Figure 3a shows PSTHs illustrating the responses to translation along eight directions in the horizontal plane (azimuthal tuning curve), whereas Fig. 3d shows PSTHs illustrating the responses to translation in eight directions in the vertical plane orthogonal to the horizontal plane and intersecting it at 45° ipsilateral azimuth (that is, the plane normal to the vector $+x+z$). Polar plots are shown in Fig. 3b, e. This neuron was maximally modulated by translation along the horizontal axis orientated at $\sim 45^\circ$ ipsilateral azimuth, but showed little modulation to translation along orthogonal axes. Translation in the direction producing a focus of expansion 135° ipsilateral azimuth produced maximal excitation whereas the opposite direction produced maximal inhibition. Using the coordinate system shown in Fig. 1a, the best axis is approximately $+x-z$.

In the vestibulocerebellum, neurons responded best to translational flow along one of three roughly orthogonal axes. Eighteen neurons responded best to translation along the y -axis. Of these, eight were classified as $+y$ neurons, and ten were classified as $-y$ neurons. The best axes of these eighteen neurons, obtained from tuning curves in the sagittal plane, are shown in Fig. 4c.

Twenty vestibulocerebellar neurons responded best to translation along horizontal axes and showed minimal modulation in response to translation along the vertical axis. Of these neurons, nine were classified as $-x-z$ neurons, as they were maximally excited by a flowfield with a focus of expansion at 135° ipsilateral azimuth. The other 11 neurons were maximally excited in response to a flowfield with a focus of expansion at 45° ipsilateral azimuth (they were

classified as $-x+z$ neurons). The best axes of these 20 neurons, obtained from azimuthal tuning curves in the horizontal plane, are shown in Fig. 4d. Figure 4e shows the best cosine fits to the mean normalized tuning curves¹³ for the four groups shown in Fig. 4c, d, where the average firing rate for each direction is expressed as a percentage of the cells' maximal firing rates. As in the case of the nBOR cells in Figs 2 and 3, the cosine fits are excellent, indicating that, for the planes tested, the response magnitude of these neurons to translation in any direction $\mathbf{d}$ was linearly related to the projection of $\mathbf{d}$ onto the neurons' preferred direction. Further, the cosine fits in Fig. 4e also show a phase shift of 90° between the $-x-z$ and $-x+z$ neurons, supporting the conclusion that the axes of the coordinate system are indeed orthogonal.

Our results indicate that the neural systems responsive to translational optic flow may be organized according to a spatial frame of reference consisting of three orthogonal axes: neurons respond best to translational optic flow along either the vertical axis (y -axis) or one of two horizontal axes orientated at 45° to the midline. It has been shown³⁻⁵ (Fig. 4a, b) that this is the spatial frame of reference for the neural systems that are responsive to rotational optic flow. The vestibular semicircular canals share this same spatial frame of reference^{3,4}, and the otolithic organs of the inner ear may also be organized according to this reference frame^{15,16}.

Therefore, the neural systems, both visual and vestibular, responsible for the encoding of self-motion may be organized in a common reference frame with six degrees of freedom, three translational and three rotational. It has been argued that the three-axis system consisting of a vertical axis and two horizontal axes orientated 45° to the midline is the most economical¹⁷, and we suggest that there are neural computational advantages of a single spatial frame of reference is used. Indeed, this is the case in other multimodal sensorimotor systems¹⁸⁻²¹. □

Methods

Anaesthesia, surgery, and extracellular recording procedures have been described^{13,14}. The translational optic flowfield stimulus was produced using a 'translator' projector suspended ~ 10 cm above the bird's head. The translator consisted of a hollow metal sphere (with a diameter of 8 cm), the surface of which was drilled with numerous small holes. A filament light source was moved along a segment of a diameter path within the sphere, so that a moving pattern of light dots was projected through the holes in the sphere's surface onto the walls, ceiling and floor of the room. Using gimbals, the axis of the spherical translational flowfield could be positioned to any orientation within three-dimensional space. The speed of the 'equatorial' dots was 1–2° per sec. For each sweep there was 5.3-s translation in one direction, followed by a 5.3-s pause, 5.3-s translation in the opposite direction, and a 5.3-s pause. PSTHs were summed from 5–10 sweeps using Spike2 software (Cambridge Electronic Designs).

Initially, once a cell was isolated, a large ($\sim 90^\circ \times 90^\circ$) handheld stimulus of random dots and lines was moved in various directions in the central areas of both visual fields. With this stimulus, neurons responded best to either vertical or horizontal motion in both hemifields. Next, using the translator, the responses of neurons to translational optic flow along the x , y and z axes were recorded. After this, tuning curves were determined by presenting translation along various axes (of 45° apart) in a particular plane. For neurons that preferred horizontal translation, azimuthal tuning curves were obtained for the horizontal plane and then elevation tuning curves were obtained for the vertical plane that intersected the horizontal plane along the axis that produced maximal modulation (Fig. 3). For neurons that preferred vertical translation, elevation tuning curves were obtained for first the sagittal plane and then the frontal plane (Fig. 2). Least-square fits of cosines to each tuning curve were obtained and the maximum was denoted as the 'best axis', within the plane of the tuning curve.

Received 7 July; accepted 22 December 1997.

1. Simpson, J. I. The accessory optic system. *Annu. Rev. Neurosci.* 7, 13–41 (1984).
2. Frost, B. J., Wylie, D. R. & Wang, Y.-C. in *Perception and Motor Control in Birds* (eds Green, P. & Davies, M.) 249–266 (Springer, Berlin, 1994).

3. Simpson, J. I., Graf, W. & Leonard, C. in *Progress in Oculomotor Research* (eds Fuchs, A. F. & Becker, W.) 475–484 (Elsevier, Amsterdam, 1981).
4. Graf, W., Simpson, J. I. & Leonard, C. S. Spatial organization of visual messages of the rabbit's cerebellar flocculus. II. Complex and simple spike responses of Purkinje cells. *J. Neurophysiol.* **60**, 2091–2121 (1988).
5. Wylie, D. R. & Frost, B. J. Responses of pigeon vestibulocerebellar neurons to optokinetic stimulation: II. The 3-dimensional reference frame of rotation neurons in the flocculus. *J. Neurophysiol.* **70**, 2647–2659 (1993).
6. Gibson, J. J. The visual perception of objective motion and subjective movement. *Psychol. Rev.* **61**, 304–314 (1954).
7. Lee, D. N. & Aronson, E. Visual proprioceptive control of standing in human infants. *Percept. Psychophys.* **15**, 529–532 (1974).
8. Owen, D. H. in *Perception & Control of Self-Motion* (eds Warren, R. & Fuchs, A. F.) 289–322 (Lawrence Erlbaum, Hillsdale, New Jersey, 1990).
9. Anderson, G. J. Perception of self-motion: psychophysical and computational approaches. *Psychol. Bull.* **99**, 52–65 (1986).
10. Krapp, H. G. & Hengstenberg, R. Estimation of self-motion by optic flow processing in single visual interneurons. *Nature* **384**, 463–466 (1996).
11. Duffy, C. J. & Wurtz, R. H. Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *J. Neurophysiol.* **65**, 1329–1345 (1991).
12. Bradley, D. C., Maxwell, M., Anderson, R. A. & Banks, M. S. Mechanisms of heading perception in primate visual cortex. *Science* **273**, 1544–1547 (1996).
13. Wylie, D. R., Kripalani, T.-K. & Frost, B. J. Responses of pigeon vestibulocerebellar neurons to optokinetic stimulation: I. Functional organization of neurons discriminating between translational and rotational visual flow. *J. Neurophysiol.* **70**, 2632–2646 (1993).
14. Wylie, D. R. & Frost, B. J. Binocular neurons in the nucleus of the basal optic root (nBOR) of the pigeon are selective for either translational or rotational visual flow. *Vis. Neurosci.* **5**, 489–495 (1990).
15. Hess, B. J. & Dieringer, N. Spatial organization of linear vestibuloocular reflexes of the rat: responses during horizontal and vertical linear acceleration. *J. Neurophysiol.* **66**, 1805–1818 (1991).
16. Macpherson, J. M. Strategies that simplify the control of quadrupedal stance. I. Forces at the ground. *J. Neurophysiol.* **60**, 204–217 (1988).
17. Simpson, J. I. & Graf, W. in *Adaptive Mechanisms in Gaze Control: Facts and Theories* (eds Berthoz, A. & Melvill-Jones, G.) 3–16 (Elsevier, Amsterdam, 1985).
18. Soechting, J. F. & Flanders, M. Moving in three-dimensional space: frames of reference, vectors, and coordinate systems. *Annu. Rev. Neurosci.* **15**, 167–191 (1992).
19. Knudsen, E. I., du Lac, S. & Esterly, S. D. Computational maps in the brain. *Annu. Rev. Neurosci.* **10**, 41–65 (1987).
20. Jay, M. F. & Sparks, D. L. Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature* **309**, 345–347 (1984).
21. Masino, T. & Knudsen, E. I. Horizontal and vertical components of head movement are controlled by distinct neural circuits in the barn owl. *Nature* **345**, 434–437 (1990).

Acknowledgements. We thank M. Dawson, D. Crewther and I. Curthoys for comments on this manuscript, and K. Lau and R. Glover for technical assistance. This research was supported by grants from the NSERC and AHFMR (to D.R.W.W.).

Correspondence and requests for material should be addressed to D.R.W.W. (e-mail: dwylie@psych.ualberta.ca).

Sniffing and smelling: separate subsystems in the human olfactory cortex

N. Sobel*, V. Prabhakaran*, J. E. Desmond†, G. H. Glover‡, R. L. Goode§, E. V. Sullivan*|| & J. D. E. Gabrieli*†

* Program in Neuroscience and Departments of † Psychology, ‡ Radiology, § ENT Surgery and || Psychiatry & Behavioral Sciences, Stanford University, Stanford, California 94305, USA
§ Palo Alto VA Hospital, Palo Alto, California 94305, USA

The sensation and perception of smell (olfaction) are largely dependent on sniffing, which is an active stage of stimulus transport and therefore an integral component of mammalian olfaction^{1,2}. Electrophysiological data obtained from study of the hedgehog, rat, rabbit, dog and monkey indicate that sniffing (whether or not an odorant is present) induces an oscillation of activity in the olfactory bulb, driving the piriform cortex in the temporal lobe, in other words, the piriform is driven by the olfactory bulb at the frequency of sniffing^{3–6}. Here we use functional magnetic resonance imaging (fMRI) that is dependent on the level of oxygen in the blood to determine whether sniffing can induce activation in the piriform of humans, and whether this activation can be differentiated from activation induced by an odorant. We find that sniffing, whether odorant is present or

absent, induces activation primarily in the piriform cortex of the temporal lobe and in the medial and posterior orbito-frontal gyri of the frontal lobe. The source of the sniff-induced activation is the somatosensory stimulation that is induced by air flow through the nostrils. In contrast, a smell, regardless of sniffing, induces activation mainly in the lateral and anterior orbito-frontal gyri of the frontal lobe. The dissociation between regions activated by olfactory exploration (sniffing) and regions activated by olfactory content (smell) shows a distinction in brain organization in terms of human olfaction.

The brains of six subjects were scanned in an experiment that contrasted the sniffing of non-odorized clean air with lack of sniffing. Sniffing induced activation primarily in the ventral temporal region (piriform, entorhinal and parahippocampal regions) and also in a small portion of the posterior and medial orbito-frontal cortex in all six subjects (Fig. 1a). Four subjects were then each rescanned four times; in each scan each subject was sniffing continuously at a different sniff-rate. Sniff-rate consistently determined the frequency of activity in the piriform cortex in all 16 scans (Fig. 2). Including the control experiments described below, sniff-induced activation occurred in the piriform of all 13 subjects tested. In 11 of the 13 subjects, sniff-induced activation was greater in the left piriform than in the right piriform (85% of subjects, $P < 0.02$).

In six additional experiments, we asked which sniffing-associated factor caused the piriform activation. First, we asked whether the sniff-induced activation was related to the motor action of sniffing or to the somatosensory stimulation induced by sniffing. Four subjects were scanned while they were sniffing with their nostrils blocked (that is, they were unsuccessfully trying to sniff); this eliminated the somatosensory stimulation associated with sniffing but maintained the motor element of sniffing. Such attempts to sniff did not induce significant activation in the piriform in any of the four subjects (Fig. 3B, d).

The same four subjects were then scanned under conditions of artificial sniffing, in which non-odorized air was puffed into the nostrils at a flow, duration, and rate similar to that of a natural sniff. This procedure, which eliminated the motor action but maintained the somatosensory element of sniffing, induced significant activation in the piriform of all four subjects (Fig. 3B, c).

An additional subject was scanned three times while sniffing with a partial occlusion of the nostrils that left 2-, 4-, or 6-mm opening. The smaller opening was associated with increased motor effort and decreased flow, whereas the larger opening was associated with decreased motor effort and increased flow. An increase in unoccluded-nostril diameter was consistently accompanied by an increase of activation in the piriform; in other words activation was related to the somatosensory sensation of air flow.

To address the possibility that sniff-induced activation may have reflected an fMRI contrast artefact (because of the periodic change in air content surrounding the nasal passages), rather than brain activity, we tested four additional subjects while they were sniffing before and after applying a topical anaesthetic to the nasal passages. Subjects were given an anaesthetic combined with a nasal dilator which, taken together, increased air flow as measured by anterior rhinometry. If sniff-induced activation were an artefact of air flow, this procedure would increase sniff-induced activation. Three of the four subjects reported a slight numbing sensation in the nostrils (but see ref. 7). The anaesthesia markedly reduced sniff-induced activation in the three subjects who reported a reduction in sensation (compare Fig. 3Ba, b), but not in the subject who did not report a reduction in sensation. The anaesthesia did not affect the perception of odours by these subjects in a standard test of olfactory identification (The University of Pennsylvania Smell Identification Test)⁸. Thus, sniff-induced activation is not an artefact related to air flow, but is instead related to brain activity induced by the sensation of air flow.

To address further the issue of possible airflow artefacts, we

scanned patient OT1, a 35-year-old woman who had undergone resection of the nerves innervating the nasal passage. She sustained a loss of olfactory and trigeminal perception in the left, but not the right, nostril. Sniffing through the intact nostril induced significant activation in the piriform, whereas sniffing through the anosmic nostril did not (compare Fig. 3Aa, b). These results show that air flow *per se* does not induce activation in the piriform. Rather, intact nerves are necessary for this activation to occur.

Another explanation for sniff-induced brain activation is that, despite the use of active-charcoal-filtered air in these experiments, the activation was odorant-induced. If the stimulus-delivery system was contaminated, piriform activations could have been induced by

an odorant rather than by sniffing. This possibility was ruled out by the second main experiment, in which we localized activation induced by olfactory content alone (smell) rather than by olfactory exploration (sniffing).

Eight subjects were repeatedly asked to determine whether any odorant was present during a scan. In these smelling tasks, the presence of odorant was alternated with the absence of odorant while sniff-rate was held constant as a baseline. Each subject was scanned twice with the pure olfactory odorants decanoic acid and vanillin. Subjects were between 89% and 100% (mean 96%) accurate in detecting odorant presence/absence. The presence of odorants induced large activations in the anterior and lateral

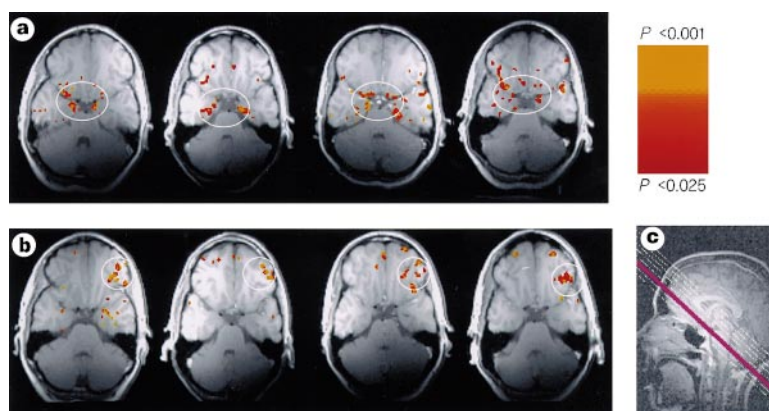


Figure 1 Brain regions activated by sniffing and smelling. **a**, The third slice (coloured purple in **c**) from four different subjects, showing sniff-induced activation primarily in the region of the piriform cortex in the temporal lobe and the medial and posterior orbito-frontal gyri in the frontal lobe (circled regions). **b**, The third slice in the same four subjects, showing activation induced by the presence of an odorant primarily in the lateral and anterior orbito-frontal gyri

(circled regions). Odorant-induced activity was greater in the right than in the left lateral orbital gyrus ($t(7) = 2.49, P < 0.05$). **c**, Slice orientation. Locations within the oblique slices were cross-referenced to standard coronal slices^{29,30}. Maximum activation induced by sniffing was at $x = -2.2$ cm, $y = 0.4$ cm, $z = -2.1$ cm, and by presence of odorant was at $x = 3$ cm, $y = 4.5$ cm, $z = -1$ cm.

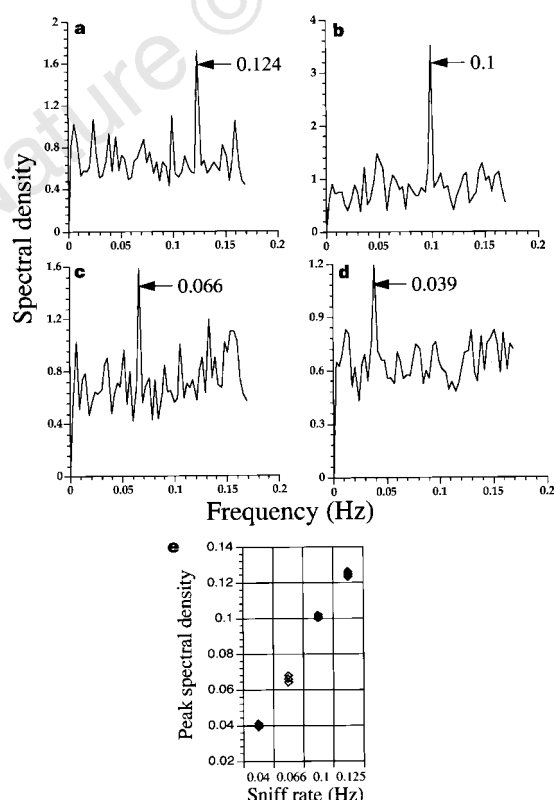


Figure 2 Fourier transform of activity within the temporal piriform during four different scans in the same subject. Subjects sniffed once every **a** 8 s (0.125 Hz), **b** 10 s (0.1 Hz), **c** 15 s (0.066 Hz) or **d** 25 s (0.04 Hz) in each scan. Sniffing frequency determined the frequency of activation in the piriform in each scan. **e**, This scatter diagram, representing the four subjects that participated in the control experiment, reveals the minimal variance in this effect. Respiration was monitored (0.38 Hz, 0.34 Hz, 0.30 Hz and 0.32 Hz in the four scans, respectively) in order to assure that overall respiration rate was not linked to sniffing rate.

orbito-frontal cortices primarily, and lesser activations in the piriform, in all 16 scans (Fig. 1b). Activity in the lateral orbito-frontal cortex was significantly greater in the right than in the left hemisphere in six of eight subjects. No activation occurred in additional 'sham' scans when no odorant was present. Thus, odorant-induced activation occurred primarily in brain regions that were different from those in which sniff-induced activation occurred, and sniff-induced activation was not related to odorant contamination of the system. (Odorant presence also induced activations in the peri-insular region, superior temporal, and various parts of the limbic system. Here, however, we focus on the dynamics of activity in the classic olfactory regions that occupy the ventral aspects of the brain.) The difference between sniff-induced and odorant-induced activation was seen in each subject simultaneously during the same scan (Fig. 4).

We have not determined which of several intranasal nerves convey the sniff-induced information. The somatosensory nature of the stimulus suggests that sniff-induced activation may be conveyed via the trigeminal nerve⁹ (CNV 5). In some species, however, the olfactory receptors and nerve (CNV 1) convey not only odorants but also mechanical pressure^{3,6}. Our results indicate that the human olfactory nerve may be mechanically sensitive. The role of sniff-induced activation in olfaction is unknown. Sniffing

may be an attentional mechanism in olfaction. A sniff may, therefore, temporarily prime the piriform for the arrival of odour information, and may thus increase the probability of detecting odours in the olfactory system. Piriform activation is driven primarily by sniff rather than smell, but odour-content information may still be found in the piriform: odour information may be temporally coded in the nervous system¹⁰. If the odour information in the piriform is temporally coded, changes in odour content could change the order of neuronal firing, without altering overall firing rate. Such a difference in temporal codes of activity would be undetectable in this fMRI analysis.

The difference in brain-activity patterns induced by sniffing and smelling corresponds broadly with the localization of the primary and secondary olfactory cortices. At present the human primary olfactory cortex is thought to lie in the piriform cortex of the temporal lobe, and the secondary olfactory cortex is believed to lie in the orbito-frontal gyri of the frontal lobe and the insula of the temporal lobe^{11–16}. This distinction concurs broadly with the structural connectivity of the olfactory system^{17–19}.

If the exploratory stage of olfactory processing (sniffing) delineates the primary olfactory cortex, the sniff-induced activations show that primary olfactory cortex extends from the ventral temporal piriform into a small region in the posterior and medial orbito-

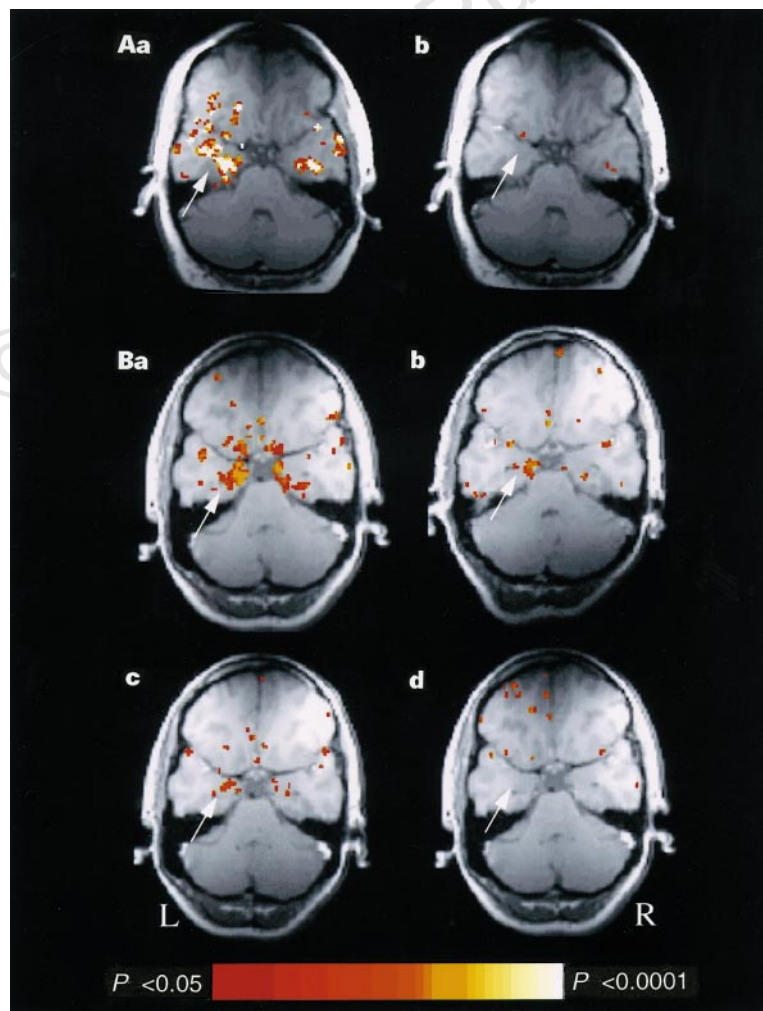


Figure 3 The component of sniffing that activates the piriform. **A,a**, Sniffing through the right, healthy, nostril in patient OT1. Piriform activation is evident. **b**, Sniffing through the left, anosmic, nostril in patient OT1. No significant activation is seen. **B,a**, Sniffing in healthy subject DP. Piriform activation is seen. **b**, Subject DP sniffing following anaesthesia. There is a reduction in activation in

comparison with **a**. **c**, Subject DP lying passively while artificial sniffing occurs. Piriform activation is seen. **d**, Subject DP performing the motor action of sniffing with the nostrils occluded (that is, there is no air flow). No significant activation is seen. Arrows point to the region of the entorhinal and temporal piriform. R, right; L, left.

frontal cortex. This continuous region of activation may correspond with the cytoarchitectural extension of the piriform into the frontal lobe in humans^{17,19}. Thus, in contrast to traditional views, primary olfactory cortex in humans may reside in both temporal and frontal lobes. If olfactory content (smelling) delineates the secondary olfactory cortex, the smell-induced activations specify the frontal-lobe portion of secondary olfactory cortex to the lateral and anterior orbito-frontal gyri (thus excluding the medial-posterior portions of the orbito-frontal cortex).

The role of sniffing in the dynamics of neural activity in the olfactory cortex may offer a new way to examine the olfactory deficits that are seen in many motor-neurodegenerative disorders^{21–23}. The olfactory deficit in diseases such as Parkinson's disease²² may be partly related to an inability to sniff rather than an inability to smell. Furthermore, the presence of both primary and secondary olfactory cortices within both frontal and temporal lobes may explain the diversity often seen in the extent and nature of olfactory deficits following temporal and frontal-lobe lesions^{19,23,24}. □

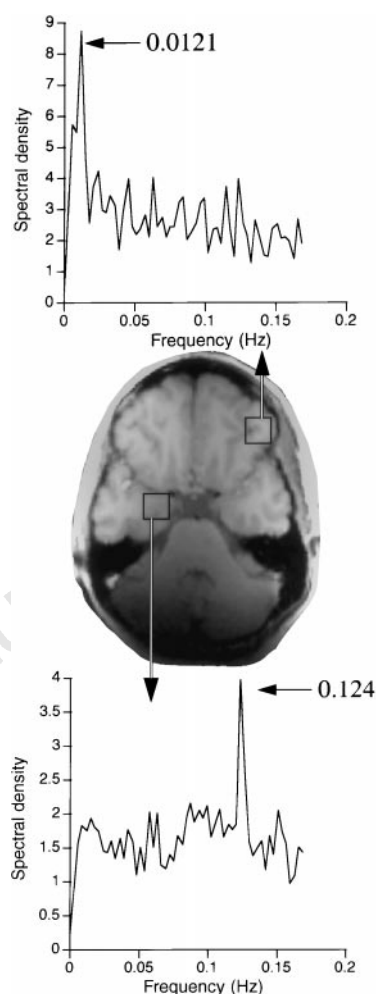


Figure 4 The dissociation between sniff-induced and odour-induced activity can be seen within the same subject at the same time (during the same scan). The subject performed a smell task (sniffing once every 8 s (0.125 Hz); odourant presence alternated with odourant absence every 40 s during the 320 s scan (0.0125 Hz)). Fourier transform of activity in the box marked in the piriform cortex (centre and bottom) and lateral orbito-frontal gyri (centre and top) reveals that whereas the primary frequency of activity in the former is determined by sniffing, activity in the latter is determined by smelling (odourant presence). In this image, the brain is seen to be performing two different tasks at two different frequencies simultaneously. Re-analysis of all the smelling tasks in all subjects at the frequency of sniffing also showed this effect.

Methods

Thirteen healthy subjects (seven women and six men, right-handed, non-smoking, mean age 28) and one unilaterally anosmic woman participated in the study. Odorant-generation methods are described in ref. 26. Odorant and no-odorant conditions were alternated, with a rise time of <500 ms, and were not associated with any auditory, visual, tactile or thermal cues.

Sniffing. During a scan, subjects viewed a screen via a mirror. An instruction was repeatedly projected to the screen every 8 s. Five repetitions of the instruction read "Sniff" and were followed by five repetitions reading "No sniff". The resulting cycle of 40 s sniffing followed by 40 s not sniffing was repeated four times, resulting in an experiment of 320 s. Subjects were instructed to maintain sniff duration while the message was projected, that is, for 1.5 s. In the somatosensory control experiments, subjects lay passively as clean air was puffed at the nostrils at 30 litres min⁻¹ for 1.5 s, at the experimental rate. Anaesthesia was performed with topical application of phenylephrine hydrochloride (0.5%) followed by xylocaine (4%).

Smelling. An instruction reading "Sniff and respond, is there an odour? Press the right button for yes or the left button for no" was projected to the screen once every 8 s throughout the entire scan. The subject sniffed and then indicated response by using the index finger of the dominant hand to press one of two buttons. The instruction was presented at a constant rate (0.125 Hz), whereas 40 s odorant presence alternated with 40 s odorant absence four times, resulting in a 320-s experiment. Thus, the difference between the alternating conditions was restricted to the presence or absence of an odorant while sniffing was maintained as a constant subtracted baseline.

fMRI acquisition. Imaging was performed using a 1.5 T MRI scanner. A bite-bar was used to prevent head motion. Eight 4-mm-thick slices with a 2-mm interslice gap were acquired. Two 5-inch-diameter local-receive coils were used. A T2*-sensitive gradient echo spiral sequence²⁷ was used, with parameters of repetition time (TR) = 720 ms, echo time (TE) = 40 ms, flip angle = 65°, in-plane resolution = 2.75 mm², number of interleaves = 4, acquisition time = 2.88 s per frame, number of frames = 115. T1-weighted flow-compensated spin-warp anatomy images (TR = 500 ms, minimum TE) were acquired as a substrate on which to overlay functional data.

fMRI analysis. Methods of fMRI analysis are described in detail in refs 26, 28. Briefly, a reference function was computed by convolving a square wave at the task frequency (frequency of sniffing in the sniffing tasks or frequency of odorant presentation in the smelling tasks) with a data-derived estimate of the haemodynamic-response function. Correlations between the reference function and the pixel response time series were computed and normalized in a statistical parametric Z map²⁸. Motion analysis was performed and corrected using an algorithm based on correlation of the image with a reference image.

Received 25 August; accepted 9 December 1997.

1. Le Magnen, J. Etude des facteurs dynamiques de l'excitation olfactive. *L'Année Psychologique* 77–89 (1945–1946).
2. Laing, D. G. Natural sniffing gives optimum odor perception for humans. *Perception* 12, 99–117 (1983).
3. Adrian, E. D. Olfactory reactions in the brain of the hedgehog. *J. Physiol.* 100, 459–473 (1942).
4. Bressler, S. L. & Freeman, W. J. Frequency analysis of olfactory system EEG in cat, rabbit, and rat. *Electro. Clin. Neurophysiol.* 50, 19–24 (1980).
5. Bressler, S. L. Relation of olfactory bulb and cortex. II. Model for driving of cortex by bulb. *Brain Res.* 409, 294–301 (1987).
6. Ueki, S. & Domino, E. F. Some evidence for a mechanical receptor in olfactory function. *J. Neurophysiol.* 24, 12–25 (1961).
7. Jones, A. S., Lancer, J. M., Shone, G. R. & Stevens, J. C. The effect of lignocaine on nasal resistance and nasal sensation of airflow. *Acta Otolaryngol.* (Stockh.) 101, 328–330.
8. Doty, R. L., Shaman, P. & Dann, M. Development of the University of Pennsylvania smell identification test: a standardized microencapsulated test of olfactory function. *Physiol. Behav.* 32, 489–502 (1984).
9. Jones, A. S., Wight, R. G., Crosher, R. & Durham, L. H. Nasal sensation of airflow following blockade of the nasal trigeminal afferents. *Clin. Otolaryngol.* 14, 285–289 (1989).
10. Wehr, M. & Laurent, G. Odour encoding by temporal sequences of firing in oscillating neural assemblies. *Nature* 384, 162–166 (1996).
11. Price, J. L. in *The Human Nervous System* (ed. Paxinos, G.) 979–1001 (Academic, San Diego, 1990).
12. Price, J. L. et al. in *Olfaction, a Model System for Computational Neuroscience* (eds Davis, J. L. & Eichenbaum, H.) 101–120 (MIT Press, Cambridge, MA, 1991).
13. Zatorre, R. J. & Jones-Gotman, M. Right-nostril advantage for discrimination of odors. *Percept. Psychophys.* 47, 526–531 (1990).
14. Zatorre, R. J. & Jones-Gotman, M. Human olfactory discrimination after unilateral frontal or temporal lobectomy. *Brain* 114, 71–84 (1991).
15. Zatorre, R. J., Jones-Gotman, M., Evans, A. C. & Meyer, E. Functional localization and lateralization of human olfactory cortex. *Nature* 360, 339–341 (1992).
16. Jones-Gotman, M. & Zatorre, R. J. Olfactory identification deficits in patients with focal cerebral excision. *Neuropsychologia* 26, 387–400 (1988).

17. Allison, A. C. The secondary olfactory areas in the human brain. *J. Anat.* **88**, 481–488 (1954).
18. Potter, H. & Nauta, W. J. H. A note on the problem of olfactory associations of the orbitofrontal cortex in the monkey. *Neuroscience* **4**, 361–367 (1979).
19. Von Bonin, G. & Green, J. R. Connections between orbital cortex and the diencephalon in the macaque. *J. Comp. Neurol.* **92**, 243–254 (1949).
20. Eslinger, P. J., Damasio, A. R. & Van Hoesen, G. W. Olfactory dysfunction in man: anatomical and behavioral aspects. *Brain Cognit.* **1**, 259–285 (1982).
21. Elian, M. Olfactory impairment in motor neuron disease: a pilot study. *J. Neurol. Neurosurg. Psychiatr.* **54**, 927–928 (1991).
22. Doty, R. L., Deems, D. A. & Stellar, S. Olfactory dysfunction in Parkinsonism: a general deficit unrelated to neurologic signs, disease stage, or disease duration. *Neurology* **38**, 1237–1244 (1988).
23. Doty, R. L. *et al.* Olfactory dysfunction in three neurodegenerative diseases. *Geriatrics* **46** suppl 1, 47–51 (1991).
24. Potter, H. & Butters, N. An assessment of olfactory deficits in patients with damage to prefrontal cortex. *Neuropsychologia* **18**, 621–628 (1980).
25. Henkin, R. I., Comiter, H., Fedio, P. & O'Doherty, D. Defects in taste and smell recognition following temporal lobectomy. *Trans. Am. Neurol. Assoc.* **102**, 146–150 (1977).
26. Sobel, N. *et al.* A method for functional magnetic resonance imaging of olfaction. *J. Neurosci. Methods* **78**, 115–121 (1997).
27. Glover, G. H. & Lai, S. Self-navigated spiral fMRI: interleaved versus single-shot. *Magn. Reson. Med.* **39** (1998).
28. Friston, K. J., Zeigler, P. & Turner, R. Analysis of functional MRI time-series. *Hum. Brain Mapp.* **1**, 153–171 (1994).
29. Talairach, J. & Tournoux, P. *Co-planar Stereotaxic Atlas of the Human Brain* (Thieme, Stuttgart, 1988).
30. Desmond, J. E. & Lim, K. O. On- and off-line Talairach registration for structural and functional MRI studies. *Hum. Brain Mapp.* **5**, 58–73 (1997).

Acknowledgements. This work was supported by the Stanford Program in Neuroscience (S.G.F.), Phil & Allen Trust, NIAAA, and NIMH. We thank L. Stryer, B. Wandell, D. Heeger, A. Pfefferbaum, G. Boynton, J. Demb, D. Peterson, G. Heit and J. Wine.

Correspondence and requests for materials should be addressed to N.S. (e-mail: nsobel@leland.stanford.edu).

Nocistatin, a peptide that blocks nociceptin action in pain transmission

Emiko Okuda-Ashitaka^{*}, Toshiaki Minami[†], Shinro Tachibana[‡], Yoshihiro Yoshihara[§], Yuji Nishiuchi[¶], Terutoshi Kimura[¶] & Seiji Ito^{*}

^{*} Department of Medical Chemistry, Kansai Medical University, Moriguchi 570, Japan

Departments of [†] Anesthesiology and [§] Biochemistry, Osaka Medical College, Takatsuki 569, Japan

[‡] School of Biological Sciences, The National University of Singapore, Singapore 119260, Singapore

[§] Department of Neuroscience, Osaka Bioscience Institute, Suita 565, Japan

[¶] Peptide Institute Inc., Protein Research Foundation, Minoh 562, Japan

Prolonged tissue damage or injury often leads to chronic pain states such that noxious stimuli evoke hyperalgesia and innocuous tactile stimuli evoke pain (allodynia)^{1,2}. The neuropeptide nociceptin^{3,4}, also known as orphanin FQ (ref. 5), is an endogenous ligand for the orphan opioid-like receptor^{6–8} which induces both hyperalgesia and allodynia when administered by injection through the theca of the spinal cord into the subarachnoid space (that is, intrathecally)^{4,9}. Here we show that the nociceptin precursor^{3,10–13} contains another biologically active peptide which we call nocistatin. Nocistatin blocks nociceptin-induced allodynia and hyperalgesia, and attenuates pain evoked by prostaglandin E₂. It is the carboxy-terminal hexapeptide of nocistatin (Glu-Gln-Lys-Gln-Leu-Gln), which is conserved in bovine, human and murine species, that possesses allodynia-blocking activity. We have also isolated endogenous nocistatin from bovine brain. Furthermore, intrathecal pretreatment with anti-nocistatin antibody decreases the threshold for nociceptin-induced allodynia. Although nocistatin does not bind to the nociceptin receptor, it binds to the membrane of mouse brain and of spinal cord with high affinity. Our results show that nocistatin is a new biologically active peptide produced from the same precursor as nociceptin and indicate that these two peptides may play opposite roles in pain transmission.

Nociceptin is processed from its precursor prepronociceptin (Fig. 1). This sequence is bounded by pairs of the basic amino acids Lys-Arg, a general cleavage site for precursor maturation. The bovine nociceptin precursor protein consists of 176 amino acid residues. It contains three additional potential cleavage sites, delineating four processing products (bPNP-2, bPNP-3, bPNP-4 and bPNP-5), as shown in Fig. 1. The prepronociceptin gene¹³ has organizational and structural features that are very similar to those encoding precursors of the endogenous opioid peptides preproenkephalin, preprodynorphin and prepro-opiomelanocortin. We therefore investigated the possibility that the nociceptin precursor might contain not only nociceptin but also an additional bioactive peptide(s) as maturation product(s).

To test whether these peptides are biologically active *in vivo*, we synthesized them chemically and investigated their effects on nociceptin-evoked pain responses in conscious mice. Simultaneous intrathecal (i.t.) injection of equal amounts of bPNP-3 and nociceptin (50 pg per mouse of each) significantly inhibited nociceptin-induced allodynia over the 50-min experiment. As shown in Fig. 2a, the allodynia caused by nociceptin was dose-dependently blocked by bPNP-3, with a half-maximal inhibitory dose (ID₅₀) (95% confidence limits) of 715 fg (34 fg–4.25 pg). However, bPNP-2, bPNP-4 and bPNP-5 did not affect allodynia at doses up to 500 pg. bPNP-3 (500 pg) did not induce allodynia by itself (not shown). Figure 2b shows the effect of synthetic peptides on nociceptin-evoked hyperalgesia by the hot-plate test. As compared with saline (17.6 ± 0.85 s, *n* = 8), nociceptin (50 pg) shortened the response latency to 8.7 ± 0.63 s, at 15 min after i.t. injection. The nociceptin-evoked hyperalgesia was significantly attenuated by 500 pg bPNP-3 and bPNP-4 (15.2 ± 1.03 s and 14.0 ± 0.31 s, respectively). The i.t. administration of bPNP-3 and bPNP-4 alone did not affect the latency period (not shown). Neither

Table 1 Inhibition of nociceptin-induced allodynia by peptides derived from bovine and mouse prepronociceptins

Species	Peptide	Sequence	ID ₅₀ (95% confidence limits) (pg)
Bovine	bPNP-3	TEPGLEEVGEIEQKQLQ	0.715 (0.034 – 4.25)
	bPNP-3-8P	EIEQKQLQ	0.125 (0.00096 – 1.33)
Mouse	mPNP-2/3	MPR-(35aa)-QLQ	1.80 (0.22 – 7.14)
	mPNP-3	AEPGADDAEEVQKQLQ	2.33 (0.25 – 15.3)
Conserved	mPNP-3-8P	EVEQKQLQ	1.07 (0.034 – 8.07)
	PNP-3-6P	EQKQLQ	14.2 (6.29 – 47.1)
	PNP-3-6PAN	QKQLQ	No inhibition at 500 pg
	PNP-3-6PAC	EQKQL	No inhibition at 500 pg

Nociceptin (50 pg per mouse) was injected simultaneously with these synthetic peptides into conscious mice. Details are described in Methods. aa, Amino acids.

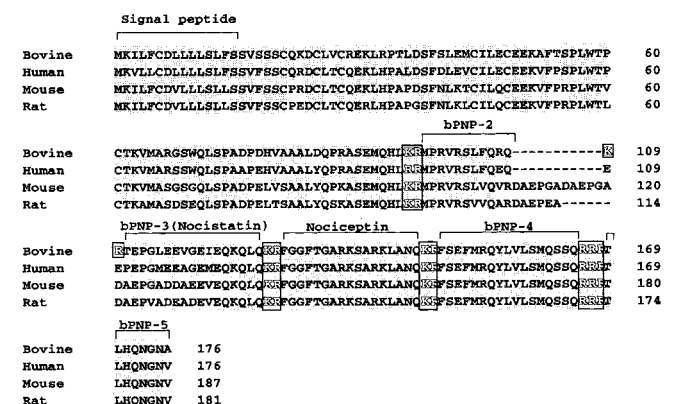


Figure 1 Alignment of deduced amino-acid sequences among bovine, human, mouse and rat prepronociceptins. Conserved amino acids are shaded and the putative proteolytic cleavage motifs are boxed. The signal peptide and putative peptides are indicated by overlines.

bPNP-2 nor bPNP-5 affected nociceptin-induced hyperalgesia. It has been reported that i.c.v. administration of NocII, the same peptide as bPNP-4, increases locomotion in mice¹⁴.

Intrathecal administration of prostaglandin (PG) E₂ induced allodynia and hyperalgesia in mice^{15–17}. As shown in Fig. 2c, d, bPNP-3 blocked PGE₂-evoked allodynia in a dose-dependent way, with an ID₅₀ (95% confidence limits) of 56.8 pg (2.49–739 pg), and reversed PGE₂-induced hyperalgesia to control levels. These results indicate that bPNP-3 may affect a common site for pain response evoked by nociceptin and PGE₂.

Although bPNP-3 is flanked by Lys-Arg pairs at residue positions 109–110 and 128–129, the human, rat and mouse precursors are devoid of the former cleavage site, which generates a processed product of length 30, 35 and 41 amino-acid residues, respectively (Fig. 1). We synthesized the following: a 41-amino-acid peptide

corresponding to residues 98–138 in the mouse precursor mPNP-2/3; a heptadecapeptide corresponding to bPNP-3 in the mouse precursor mPNP-3; C-terminal octapeptides from bPNP-3 (bPNP-3-8P) and from mPNP-3 (mPNP-3-8P); and the hexapeptide Glu-Gln-Lys-Gln-Leu-Gln (PNP-3-6P), which is 100% conserved among species. As shown in Table 1, all these synthetic peptides suppressed nociceptin-induced allodynia, but the pentapeptides Glu-Gln-Lys-Gln-Leu (PNP-3-6PΔC) and Glu-Lys-Gln-Leu-Gln (PNP-3-6PΔN) at 500 pg did not affect it. These results show that these processed parts of the nociceptin precursors have similar biological activities in different species and that the C-terminal consensus portion is crucial for activity. Structural and conformational constraints for nociceptin activation are most stringent in the amino-terminal part (residues 1–5, Phe-Gly-Gly-Phe-Thr) of the nociceptin molecule¹⁸, but the C-terminal part

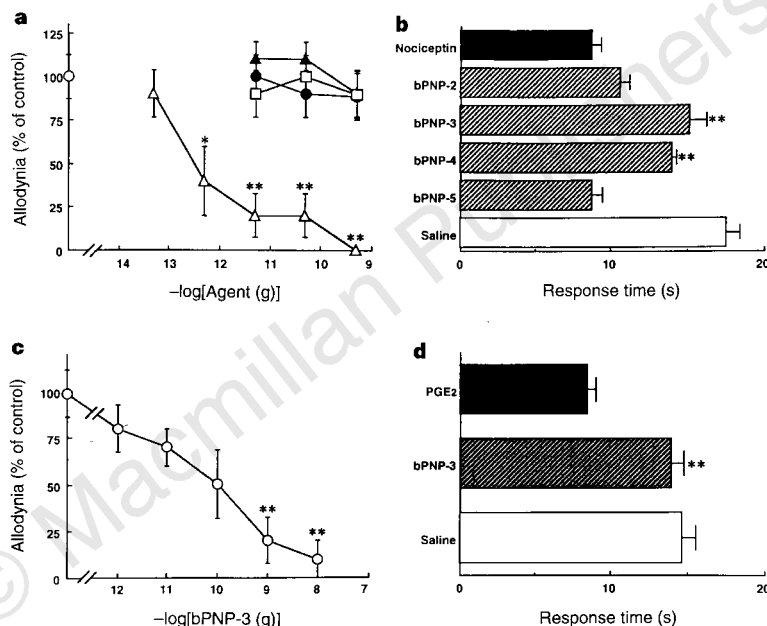


Figure 2 Effects of putative peptides deduced from bovine prepronociceptin on allodynia (**a, c**) and hyperalgesia (**b, d**). Nociceptin (**a**, 50 pg per mouse) or PGE₂ (**c**, 10 ng per mouse) was simultaneously injected i.t. with various doses of bPNP-2 (●), bPNP-3 (Δ in **a**, ○ in **c**), bPNP-4 (▲) or bPNP-5 (□). **b**, Nociceptin (50 pg) was

injected without or with 500 pg of the indicated peptide. **d**, PGE₂ (10 ng) was injected without or with 10 ng of bPNP-3. Control mice were given physiological saline. Each point represents the mean ± s.e.m. (*n*, 6–10). **P* < 0.05, ***P* < 0.01, versus the nociceptin-injected (**a, b**), and the PGE₂-injected (**c, d**) group.

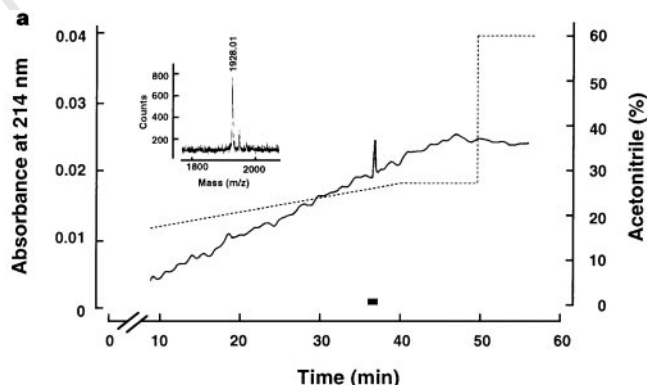
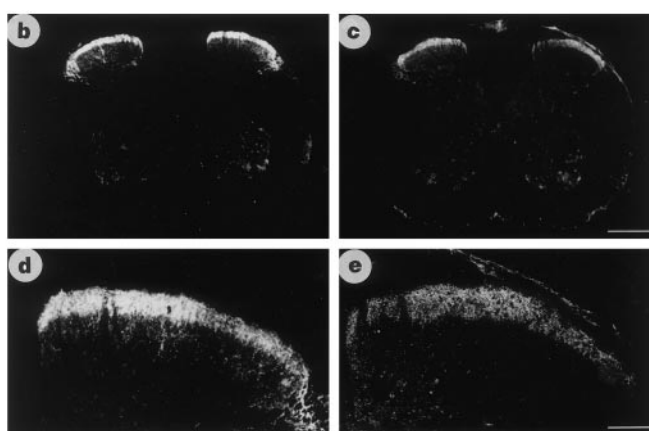


Figure 3 Existence of PNP-3. **a**, Elution profile of endogenous bPNP-3 from a TSK-ODS column. The eluate was monitored by its absorbance at 214 nm. The immunoreactive fraction against anti-bPNP-3 antibody is indicated by a black bar. Inset shows a molecular mass spectrometry profile of the immunoreactive fraction from the TSK-ODS column. **b–e**, Immunofluorescent confocal images of



mouse spinal cord with anti-mPNP-3 or anti-nociceptin antiserum. A transverse section of mouse spinal cord was incubated with anti-mPNP-3 (**b, d**) or anti-nociceptin (**c, e**) antibody followed by Cy3-conjugated anti-rabbit IgG. Scale bar, 250 μm (**b, c**); 100 μm (**d, e**).

(residues 12–17, Glu-Gln-Lys-Gln-Leu-Gln) appears to be the minimal essential core in the PNP-3 sequence.

To confirm the existence of the new peptide *in vivo*, we attempted to purify endogenous bPNP-3. Crude extract from bovine brains (203 g) was subjected to size-exclusion chromatography on Bio-Gel P-4, followed by anti-bPNP-3 immunoaffinity chromatography and reverse-phase high-performance liquid chromatography (RP-HPLC). The RP-HPLC profile (Fig. 3a) showed a single peak of absorbance at 214 nm (A_{214}) which eluted after 36.9 min, coinciding with the immunoreactivity detected by anti-bPNP-3 antibody. The sequence of the immunoreactive fraction was Thr-Glu-Pro-Gly-Leu-Glu-Glu-Val-Gly-Glu-Ile-Glu-Gln-Lys-Gln-Leu-Gln¹⁷, which was identical to that of bPNP-3 derived from bovine prepronociceptin. The identity of this highly purified peptide (yield ~300 pmol after three chromatography steps; relative molecular mass 1928.01) was confirmed by mass spectrometry (Fig. 3a, inset). A computer search (PRF/SEQDB data base, Protein Research Foundation, Osaka, Japan) revealed no homology to other known peptides.

We examined the distribution of the new peptide in mouse spinal cord by using anti-mPNP-3 antibodies. mPNP-3-like immunoreactivity was most abundant in the superficial laminae of the spinal dorsal horn (Fig. 3b, d). Addition of mPNP-3 peptide completely abolished specific staining (not shown); furthermore, the distribution of mPNP-3-like immunoreactivity appeared to be almost identical to that of nociceptin detected by using anti-nociceptin antibody (Fig. 3c, e).

To determine whether the endogenous mPNP-3-immunoreactive peptide participates in pain transmission in the spinal cord, we tested the effect of an anti-mPNP-3 antibody on nociceptin-induced allodynia. Groups of mice were injected i.t. with rabbit anti-mPNP-3 IgG or normal rabbit IgG one hour before i.t. injection of nociceptin. The dose dependence of nociceptin-induced allodynia in mice after no pretreatment showed a bell-shaped pattern in the dose range 1–100 pg per mouse. The dose–response curve was shifted to the left by 2.5 orders at the half-maximal effective dose (ED_{50}) as a result of pretreatment of mice i.t. with antibody against mPNP-3, but not with normal IgG (Fig. 4a). This result indicates that endogenous mPNP-3-immunoreactive peptide, possibly mPNP-2/3, inhibits nociceptin-evoked allodynia in the spinal cord.

bPNP-3 did not displace bound [³H]nociceptin nor did it prevent the inhibition by nociceptin of forskolin-induced accumulation of cyclic AMP in Chinese hamster ovary cells transfected with the nociceptin receptor (results not shown). To investigate whether

bPNP-3 works through a specific receptor, we used an [¹²⁵I]-labelled ligand produced from synthetic peptide analogues containing tyrosine at various positions. Lys¹⁴(Tyr)-bPNP-3, an analogue with a tyrosine residue on the ϵ -amino group of Lys 14 of bPNP-3, blocked nociceptin-induced allodynia with an ID_{50} of 0.154 pg, equivalent to bPNP-3. As shown in Fig. 4b, [¹²⁵I]Lys¹⁴(Tyr)-bPNP-3 specifically bound to the crude membrane prepared from mouse brain in a saturable manner. Scatchard analysis yielded a dissociation constant of 5 nM and a maximal binding capacity (B_{max}) of 175 fmol per mg protein. This value of B_{max} is comparable to those found for the μ - and δ -opioid receptors on mouse brain¹⁹. The natural ligand bPNP-3 displaced [¹²⁵I]Lys¹⁴(Tyr)-bPNP-3 bound to brain membrane in a manner similar to unlabelled Lys¹⁴(Tyr)-bPNP-3 (results not shown). Results were similar for mouse spinal cord. This binding study confirmed that bPNP-3 binds to the membrane of mouse brain and spinal cord with high affinity.

We have demonstrated the existence and biological activity *in vivo* of peptide bPNP-3, which we call nocistatin, derived from prepronociceptin. Preproenkephalin, preprodynorphin and preproopioidmelanocortin all contain several different repetitive peptides with related or distinct functions; we have now identified two biologically active peptides, nociceptin and nocistatin, which are generated from the same precursor and play opposite roles in pain transmission. Pharmacological opioid drugs have powerful analgesic effects in mammals. The discovery of nocistatin may lead to the design of a novel analgesic devoid of the addiction and dependence associated with morphine. □

Methods

cDNA cloning of bovine prepronociceptin. A cDNA library of bovine hippocampus and entorhinal cortex was primed with oligo(dT)₁₇ primer and size-selected from 1 to 2.3 kb, and constructed in a λ ZAPII vector. The library consisting of 8.8×10^5 phages was hybridized with a cDNA fragment (R1) from PCR containing the rat prepronociceptin sequence (residues 225–490)⁴. Hybridization was done as described²⁰. Three positive phage clones were isolated, plasmid DNAs were rescued from them using ExAssist helper phage (Stratagene) and sequenced. The longest clone, containing a 1,209-bp cDNA insert, encodes a protein of 176 amino acids, and shares 86, 74 or 72% overall amino-acid homology with human, mouse or rat prepronociceptins, respectively.

Behavioural analysis. Male ddY mice weighing 22 ± 2 g were used. Drugs (in vehicle; 5 μ l) were injected slowly into the subarachnoid space between the L5 and L6 vertebrae. Subsequent behaviour was analysed as described^{16,17}. Mechanical allodynia was assessed every 5 min over a 50-min period after i.t. injection, by lightly stroking the flanks of the mice with a paintbrush. The effects of peptides on nociceptin- and PGE₂-induced allodynia were assessed at the point of maximal allodynic effect ($83.3 \pm 10.5\%$) at 10 and 5 min after i.t. injection, respectively. For hyperalgesia, mice were placed on a hot plate maintained at 55 °C, and the elapsed time until the mice showed the first avoidance response was measured 15 and 30 min after i.t. injection of nociceptin and PGE₂, respectively.

Antibodies against bPNP-3 and mPNP-3. Anti-bPNP-3 and anti-mPNP-3 polyclonal antisera were made in rabbits with synthetic peptides conjugated to keyhole limpet haemocyanin as immunogen, as described²⁰. A peptide-specific antibody was purified from anti-bPNP-3 antiserum by using a bPNP-3-8P-conjugated Sepharose–4B column. IgG purified from anti-mPNP-3 antisera by protein G–Sepharose (Pharmacia) was used for behavioural analysis. Antisera against mPNP-3 and nociceptin (Phoenix) were specific for the cognate peptide and did not crossreact with dynorphin A(1–13), β -endorphin, or Leu- and Met-enkephalin by radioimmunoassay. These antisera did not crossreact with crude extracts prepared from NS20Y cells (donated by A. Ogura) treated with dibutyryl cAMP²¹ and COS-7 cells transfected with mouse prepronociceptin cDNA by western blot analysis, suggesting that antisera against mPNP-3 and nociceptin may not recognize prepronociceptin.

Purification of endogenous bPNP-3 from bovine brain. A crude extract⁴ prepared from hippocampus, hypothalamus, entorhinal cortex and septum of

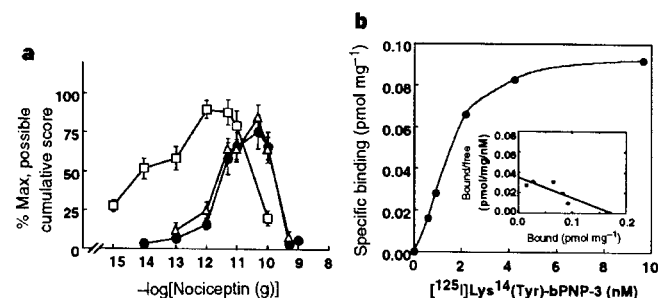


Figure 4 Biological action of PNP-3. **a**, Effect of i.t. pretreatment of anti-mPNP-3 antibody on nociceptin-induced allodynia. Mice were not pretreated (●) or were pretreated i.t. with 10 μ g of anti-mPNP-3 IgG (□) and 10 μ g of normal rabbit IgG (△) 1 h before i.t. injection of various doses of nociceptin. The allodynic score (mean $\pm$ s.e.m., $n = 6$) is expressed as a percentage of the maximum possible cumulative score over the 50-min period. **b**, Specific binding of [¹²⁵I]Lys¹⁴(Tyr)-bPNP-3 to the membrane of mouse brain. The Scatchard plot (inset) was transformed from the values of specific [¹²⁵I]Lys¹⁴(Tyr)-bPNP-3 binding. The results are the mean of triplicate determinations.

bovine brain (203 g) was lyophilized, dissolved in 5 ml 0.1 M acetic acid and applied to a Bio-Gel P-4 column (1.5 × 100 cm) equilibrated with 0.1 M acetic acid. Fractions of $M_r \sim 1,900$, taken around the eluates containing nociceptin (M_r 1,810), as determined using a cAMP-inhibition assay⁴, were pooled, neutralized and subjected to anti-bPNP-3 antibody-conjugated affinity chromatography. The column (0.7 × 3 cm) was washed with 50 ml 0.1 M Tris-Cl containing 100 mM NaCl and the adsorbed material was eluted with 0.2 M glycine-HCl, pH 2.5. Eluates containing bPNP-3-immunoreactive material were combined, subjected to RP-HPLC using a TSK ODS-120T column (4.6 × 250 mm, Tosoh, Tokyo) and eluted with a 40-min linear gradient of 15–27.5% CH₃CN in 0.1% trifluoroacetic acid. The immunoreactive fraction was sequenced using a gas-phase automatic protein sequencer (type PSQ-1; Shimadzu) and mass spectrometry on a MALDI-TOF mass type Voyager Elite (Perspective).

Immunohistochemistry. A ddY mouse was anaesthetized with pentobarbital and perfused with PBS, followed by 4% paraformaldehyde in phosphate buffer. The spinal cord was removed, postfixed overnight at 4 °C in the same fixative, and kept in cold phosphate buffer containing 30% sucrose. Frozen sections (40 μm) were cut with a microtome, incubated with antisera against mPNP-3 (at a dilution of 1:8,000) or nociceptin (1:2,000) overnight at room temperature, washed with PBS containing 10% normal goat serum and 0.2% Triton X-100, and processed with Cy3-anti-rabbit IgG (Jackson Immuno-Research). Digital images were captured on a BioRad MRC-600 confocal microscope equipped with a krypton/argon laser.

Binding assay. [¹²⁵I]Lys¹⁴(Tyr)-bPNP-3 (3.7 TBq mmol⁻¹) was prepared by the chloramine-T method²² and purified on a Bio-Gel P-2 column. For binding assay, the P2 fraction of a mouse brain or spinal cord (100 μg protein) was incubated with various concentrations of [¹²⁵I]Lys¹⁴(Tyr)-bPNP-3 in 200 μl 50 mM Tris-Cl (pH 8.5) containing 0.1 M NaCl, 10 mM CaCl₂, 10 mM MgCl₂, 1 mM EDTA, 100 μM phenylmethylsulphonyl fluoride, 0.1 mg ml⁻¹ bacitracin, 0.1 mg ml⁻¹ leupeptin, and 1 mg ml⁻¹ BSA at 4 °C for 1 h and the radioactivity associated with a Whatman GF/C filter was counted by a Beckman γ-radiation counter. Nonspecific binding was determined using a 1,000-fold excess of unlabelled Lys¹⁴(Tyr)-bPNP-3 in the reaction mixture. The specific binding was calculated by subtracting the nonspecific binding from the total binding.

Received 4 July; accepted 8 December 1997.

1. Woolf, C. J. in *Textbook of Pain* (eds Wall, P. D. & Melzack, R.) 101–112 (Churchill Livingstone, Edinburgh, 1994).
2. Dray, A., Urban, L. & Dickenson, A. Pharmacology of chronic pain. *Trends Pharmacol. Sci.* 15, 190–197 (1994).
3. Meunier, J. C. *et al.* Isolation and structure of the endogenous agonist of opioid receptor-like ORL1 receptor. *Nature* 377, 532–535 (1995).
4. Okuda-Ashitaka, E. *et al.* Identification and characterization of an endogenous ligand for opioid receptor homologue ROR-C: its involvement in allodynic response to innocuous stimulus. *Mol. Brain Res.* 43, 96–104 (1996).
5. Reinscheid, R. K. Orphanin FQ: a neuropeptide that activates an opioidlike G protein-coupled receptor. *Science* 270, 792–794 (1995).
6. Mollereau, C. *et al.* ORL1, a novel member of the opioid receptor family: cloning, functional expression and localization. *FEBS Lett* 341, 33–38 (1994).
7. Bunzow, J. R. *et al.* Molecular cloning and tissue distribution of a putative member of the rat opioid receptor gene family that is not a μ , δ or κ opioid receptor type. *FEBS Lett.* 347, 284–288 (1994).
8. Fukuda, K. *et al.* cDNA cloning and regional distribution of a novel member of the opioid receptor family. *FEBS Lett.* 343, 42–46 (1994).
9. Hara, N. *et al.* Characterization of nociceptin hyperalgesia and allodynia in conscious mice. *Br. J. Pharmacol.* 121, 401–408 (1997).
10. Notheracker, H. P. *et al.* Primary structure and tissue distribution of the orphanin FQ precursor. *Proc. Natl Acad. Sci. USA* 93, 8677–8682 (1996).
11. Pan, Y. X., Xu, J. & Pasternak, G. W. Cloning and expression of a cDNA encoding a mouse brain orphanin FQ/nociceptin precursor. *Biochem. J.* 315, 11–13 (1996).
12. Houtani, T., Nishi, M., Takeshima, H., Nukada, T. & Sugimoto, T. Structure and regional distribution of nociceptin/orphanin FQ receptor. *Biochem. Biophys. Res. Commun.* 219, 714–719 (1996).
13. Mollereau, C. *et al.* Structure, tissue distribution, and chromosomal localization of the prepronociceptin gene. *Proc. Natl Acad. Sci. USA* 93, 8666–8670 (1996).
14. Florin, S., Suaudeau, C., Meunier, J.-C. & Costentin, J. Orphan neuropeptide NocII, a putative pronociceptin maturation product, stimulates locomotion in mice. *NeuroReport* 8, 705–707 (1997).
15. Taiwo, Y. O. & Levine, J. D. Prostaglandins inhibit and endogenous pain control mechanism by blocking transmission at spinal noradrenergic synapses. *J. Neurosci.* 8, 1346–1349 (1988).
16. Minami, T. *et al.* Allodynia evoked by intrathecal administration of prostaglandin E₂ to conscious mice. *Pain* 57, 217–223 (1994).
17. Minami, T. *et al.* Characterization of EP-receptor subtypes involved in allodynia and hyperalgesia induced by intrathecal administration of prostaglandin E₂ to mice. *Br. J. Pharmacol.* 112, 735–740 (1994).
18. Reinscheid, R. K., Ardati, A., Monsma, F. J. Jr & Civelli, O. Structure–activity relationship studies on the novel neuropeptide orphanin FQ. *J. Biol. Chem.* 271, 14163–14168 (1996).
19. Matthes, H. W. D. *et al.* Loss of morphine-induced analgesia, reward effect and withdrawal symptoms in mice lacking the μ -opioid-receptor gene. *Nature* 383, 819–823 (1996).
20. Okuda-Ashitaka, E. *et al.* Suppression of prostaglandin E receptor signaling by the variant form of EP₁ subtype. *J. Biol. Chem.* 271, 31255–31261 (1996).

21. Saito, Y. *et al.* Molecular cloning and characterization of a novel form of neuropeptide gene as a developmentally regulated molecule. *J. Biol. Chem.* 271, 15615–15622 (1996).
22. Amano, F., Gottesman, M. M. & Pastan, I. Epidermal growth factor-dependent growth of human KB cells in a defined medium and altered growth factor requirements of KB mutants resistant to EGF–*Pseudomonas* exotoxin conjugates. *J. Cell. Physiol.* 135, 502–508 (1988).

Acknowledgements. We thank K. Uegaki and N. Yumoto for peptide synthesis; H. Minakata for amino-acid sequence analysis and mass spectrometry; and Y. Masu for discussion and comments. This work was supported by the Ministry of Education, Science, Culture, and Sports of Japan and the Japan Private School Promotion Foundation.

Correspondence and requests for materials should be addressed to S.I. (e-mail: ito@takii.kmu.ac.jp). The nucleotide sequence of nocistatin has been deposited in the GenBank/EMBL/DBJ databases under accession number AB005251.

Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito

O. Billker*, V. Lindo†, M. Panico†, A. E. Etienne†, T. Paxton†, A. Dell†, M. Rogers‡, R. E. Sinden* & H. R. Morris†

* Department of Biology, Imperial College, London SW7 2BB, UK

† Department of Biochemistry, Imperial College, London SW7 2AZ, UK

‡ M-Scan Inc., 606 Brandywine Parkway, West Chester, Pennsylvania 19380, USA

Malaria is transmitted from vertebrate host to mosquito vector by mature sexual blood-living stages called gametocytes^{1,2}. Within seconds of ingestion into the mosquito bloodmeal, gametocytes undergo gametogenesis. Induction requires the simultaneous exposure to at least two stimuli *in vitro*: a drop in bloodmeal temperature to 5 °C below that of the vertebrate host^{1–3}, and a rise in pH from 7.4 to 8.0–8.2 (refs 1, 4). *In vivo* the mosquito bloodmeal has a pH of between 7.5 and 7.6 (refs 5, 6). It is thought that *in vivo* the second inducer is an unknown mosquito-derived gametocyte-activating factor^{5,7,8}. Here we show that this factor is xanthurenic acid. We also show that low concentrations of xanthurenic acid can act together with pH to induce gametogenesis *in vitro*. Structurally related compounds are at least ninefold less effective at inducing gametogenesis *in vitro*. In *Drosophila* mutants with lesions in the kynurenine pathway of tryptophan metabolism (of which xanthurenic acid is a side product), no alternative active compound was detected in crude insect homogenates. These data could form the basis of the rational development of new methods of interrupting the transmission of malaria using drugs or new refractory mosquito genotypes to block parasite gametogenesis.

Mature malaria gametocytes circulate in the bloodstream of the vertebrate host, arrested at the G₀ phase of the cell cycle. Gametocytes undergo gametogenesis within 15 s of being exposed to a decrease in temperature combined with an increase in pH, or the addition of a heat-stable, low-molecular-weight gametocyte activating factor (GAF)^{2,7}. Within 10 min both male and female gametocytes have escaped from the enveloping erythrocytes, and the male has produced eight flagellate microgametes in a process termed exflagellation. To characterize GAF we have prepared active extracts from homogenates of 100–150 *Anopheles stephensi* pupae by boiling, methanol-chloroform extraction and ultrafiltration, followed by purification by reverse-phase high-performance liquid chromatography (HPLC) on a C-18 column eluted with acetic acid and an *n*-propanol gradient. GAF activity across the column fractions was monitored by an *in vitro* bioassay for exflagellation using *Plasmodium berghei*⁸. A single peak of biological activity was detected, corresponding to a ultraviolet-active component eluting at 15.8 min (Fig. 1). These data were compared with equivalent mass-spectrometric (MS) screening data to cross-correlate mass with the active fractions obtained. Mass spectrometric

screening was performed by electrospray quadrupole orthogonal acceleration time of flight (Q-TOF) mass spectrometry (ES-MS) and MS-MS technology⁹⁻¹¹, confirmed in the MS mode by matrix-assisted laser desorption (MALDI TOF MS) using a Voyager Elite mass spectrometer. Mass candidates for GAF at this stage of purification included signals at mass:charge ratios (m/z) of 206 and 377. The bioassay-positive GAF fraction from HPLC was further subjected to high-voltage (3 kV) paper electrophoresis (HVPE) in pyridine/acetic buffer at pH 6.5, again followed by bioassay and mass spectrometric ES-MS and MALDI screening, for which 1 cm strips cut from the paper in neutral, anodic and cathodic directions from the origin were eluted with 5% acetic acid. Biological activity was found at a relative mobility of 0.4 with

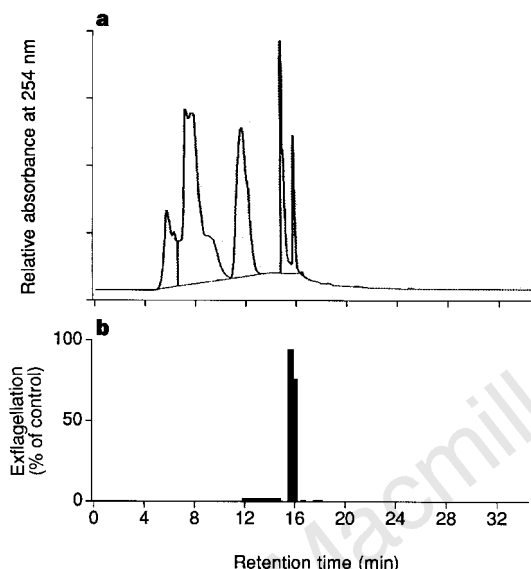


Figure 1 Distribution of GAF activity following fractionation of crude extracts of *Anopheles stephensi* pupae by reverse-phase HPLC. **a**, Absorbance at 254 nm. **b**, Bioassay of GAF activity measured by induction of exflagellation of *P. berghei* *in vitro* and expressed as a percentage of that induced by a reference solution (RPMI 1640 with 20 mM HEPES, pH 8.0).

respect to aspartic acid, showing that GAF is a small acidic substance carrying a net negative charge (derived from carboxyl, phosphate or sulphate group(s)) at neutral pH. Mass-spectrometric analysis of the corresponding fractions discounted m/z 377 and other signals as being related to GAF, leaving m/z 206 as the most likely candidate for being the GAF quasimolecular ion. Further HVPE experiments were then performed to distinguish between the anionic functionalities listed above, and to examine the possibility of cationic groups, such as primary amino groups, in the molecule. The HVPE/bioassay/mass-spectrometric screening protocol at pH 2.1 showed the GAF to be neutral, suggesting that a carboxyl group was the source of ionic behaviour and pre-empting the likelihood of a primary amino group in the molecule. The Q-TOF ES-MS spectrum associated with the biologically active fraction now showed an intense signal at m/z 206 ($M + H$)⁺, where M is the molecule, confirming that the compound has a mass of 205, thereby indicating that it has an odd number of nitrogen atoms in the structure. A stronger signal at m/z 204 ($M - H$)⁻ for GAF run in the negative-ion mode confirmed the anionic behaviour seen in the HVPE experiments. Another study has recently suggested that GAF has a mass of either 205 or 219, although further structural conclusions could not be drawn¹².

Here we combine data from the above experiments with micro-chemical derivatization studies¹³⁻¹⁵, high-resolution accurate mass measurement of the 204 ($M - H$)⁻ ion and tandem MS-MS analysis on the Q-TOF instrument to determine the structure of GAF. Derivative formation of GAF by reaction with acetic anhydride in pyridine gave a mass shift of 42, indicating the incorporation of one acetyl group under the reaction conditions used. The relative intensities of positive-ion (very weak) and negative-ion (intense) data suggest the incorporation of the acetyl group on a secondary nitrogen rather than hydroxyl function. Esterification of GAF in methanolic HCl showed a mass shift (of 14) to m/z 220 ($M + H$)⁺, indicating the presence of one carboxyl group, which, in contrast to the α -carboxyl of an amino acid, for example, was not fully blocked in the reaction time used (1 h). This suggests an $\alpha\beta$ unsaturated carboxyl of the type seen in some pyrrole and pyridino compounds¹⁶. Diazomethane treatment, which would methylate carboxyl and acidic hydroxyls in a molecule, showed up to three such groups (one already shown to be carboxyl) with a shift to m/z 248 in the positive-ion mass spectrum.

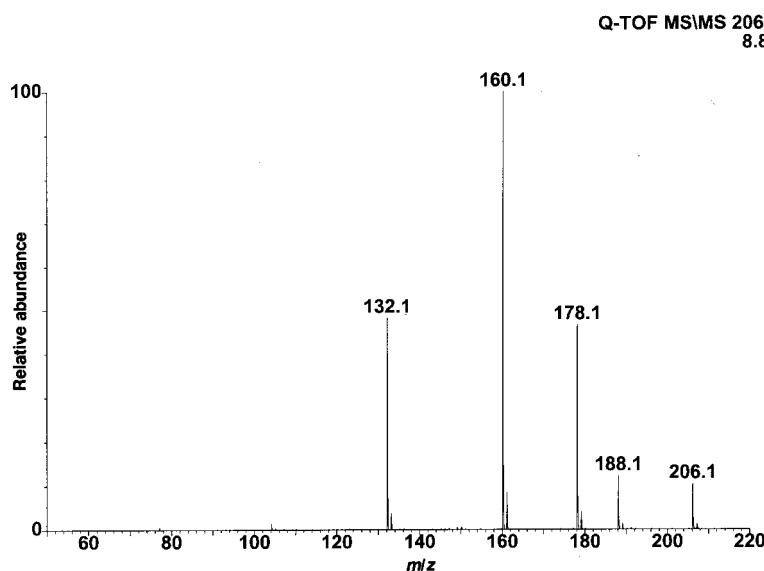


Figure 2 MS-MS collisionally activated decomposition spectrum of the GAF quasimolecular ion m/z 206 at 25–30 eV in argon on the Q-TOF instrument. The base peak at m/z 160 corresponds to loss of the elements of water and carbon monoxide to give a stable ring-contracted hydroxyindole acylium ion. When the

cone voltage-induced fragment ion at m/z 160 is subjected to MS-MS analysis, m/z 178 (water addition) was observed together with the low-mass fragment ions, which indicated that m/z 188 and 160 alone are the important primary fragments.

Despite the small amount of purified GAF available for study, accurate mass measurement of the quasimolecular ion of GAF was achieved in the negative-ion mode using fast atom bombardment (FAB) mass spectrometry. The accurate mass ($M - H$)⁻ obtained was 204.0290, suggesting that GAF has an atomic composition C₁₀H₇NO₄. This atomic composition is significant in having a very large double-bond equivalent number of eight, indicating that GAF has a highly unsaturated aromatic or pseudoaromatic structure. This finding, together with the functional groupings already suggested from the HVPE and microchemical derivative-forming experiments, means that there are very few ways of arranging such data into a probable molecular structure. In considering the alternatives, the Q-TOF collisionally activated decomposition MS-MS data on the quasi-molecular ion at 206 (shown in Fig. 2) are important. The spectrum is deceptively simple, indicating losses of water and carbon monoxide (see legend to Fig. 2) with the major fragment ion at m/z 160 corresponding at first sight to either loss of formic acid or the combination of its component parts (water and carbon monoxide). At moderate collision energies, m/z 160 is the base peak, suggesting a very stable structure of the aromatic type, which can decompose only at higher energies to give m/z 132, 104, 77 and 51. These data indicate a combination of the pyridino structure discussed earlier with an aryl unit to give a highly unsaturated core structure with one carboxyl group; the remaining elements of the atomic composition are assigned as two hydroxyl functional groups. Such a compound, xanthurenic acid (Fig. 3) has been described as a product of tryptophan metabolism. Figure 3 shows the structure along with some possible tautomers and resonance structures relevant to its chemical and fragmentation behaviour.

Synthetic xanthurenic acid co-chromatographs on HPLC with GAF and possesses identical MS, MS-MS and derivative-forming

properties. Under standard assay conditions at pH 7.4, synthetic xanthurenic acid induces exflagellation in *P. berghei* with an effector concentration for half-maximal response (EC₅₀) of 9 μ M (Fig. 4a). Equivalent concentrations of the molecule identified as GAF were found in extracts from *A. stephensi* adults. Xanthurenic acid is indistinguishable from GAF by the criteria used, and we therefore conclude that it is GAF.

Xanthurenic acid has been described as a lateral reaction product of the ommochrome pathway of eye-pigment synthesis in insects¹⁷ (Fig. 4b), including the malaria vector *A. gambiae*¹⁸. Our identification of GAF as xanthurenic acid therefore provides an explanation for the observed GAF activity in mosquito heads⁷. In *Drosophila*, xanthurenic acid is deposited in Malpighian tubules and excreted in the faeces¹⁹. Among the metabolites that are structurally related to xanthurenic acid, kynurenic acid and quinaldic acid were much less active (EC₅₀ of 180 and 80 μ M, respectively) whereas the metabolic precursors kynurenine and 3-hydroxykynurenine showed no activity over the concentration range tested (Fig. 4a), suggesting that the strict structural requirements for GAF activity are found only in xanthurenic acid.

Further evidence that the ommochrome pathway is the source of GAF comes from the use of *D. melanogaster* eye-colour mutants (Table 1). GAF activity was present in homogenates of wild-type flies, but not in *vermillion* and *cinnabar* mutants, which lack xanthurenic acid as a result of separate enzyme defects upstream of 3-hydroxykynurenine²⁰ (Fig. 4b). GAF activity was further absent from *white* flies, in which the lesion affects the cellular uptake of 3-hydroxykynurenine precursors and pteridine precursors²¹, but activity was present in *brown* mutants, in which only the pteridine pathway of pigment synthesis is affected²².

Despite its ability to initiate the mosquito phase of the malaria cycle *in vitro*, xanthurenic acid is not specific to mosquitoes, and is

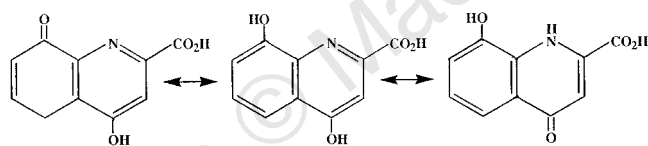


Figure 3 Some resonance structures of xanthurenic acid that are relevant to its pH-related chemistry and fragmentation behaviour.

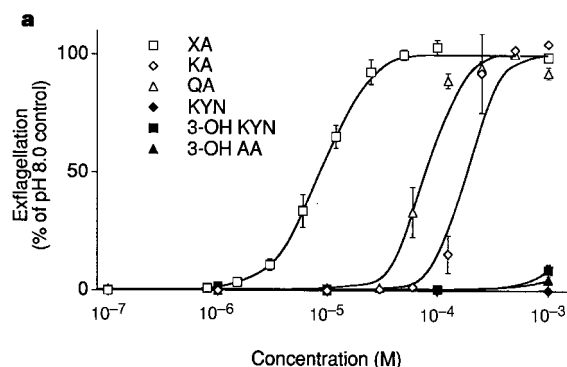


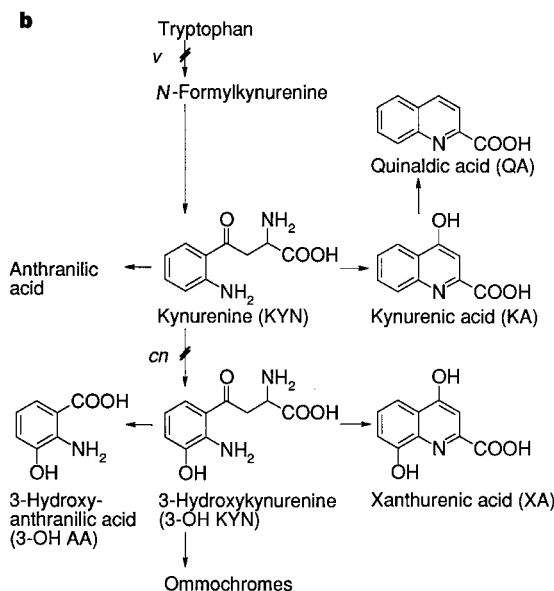
Figure 4 Ommochrome pathway of tryptophan metabolism and the effect of metabolites on exflagellation. **a**, Dose-response curves for the induction of exflagellation by xanthurenic acid and related metabolites at an otherwise non-permissive pH of 7.4. See **b** for explanation of abbreviations. The exflagellation response to medium at pH 8.0 served as the 100% reference value. **b**, Ommochrome pathway of tryptophan metabolism in insects (modified from ref. 26) showing structures for metabolites tested for GAF activity and metabolic blocks in the *Drosophila* mutants *vermillion* (v) and *cinnabar* (cn).

Table 1 GAF activity in *Drosophila melanogaster* eye-colour mutants

Mutant	Exflagellation response*
Wild type (Canton-S)	83.2 $\pm$ 14.5
<i>vermillion</i>	0.8 $\pm$ 0.7
<i>cinnabar</i>	0.1 $\pm$ 0.2
<i>brown</i>	74.0 $\pm$ 2.6
<i>white</i>	0.6 $\pm$ 0.5
Medium only (negative control)	0.2 $\pm$ 0.4

Supernatants from homogenates of 20 flies in medium (RPMI 1640, 20 mM HEPES) were adjusted to pH 7.5 before testing.

* Exflagellation responses are mean activity $\pm$ s.d. of 3 homogenates, expressed as a percentage of a 100 μ M xanthurenic acid control.



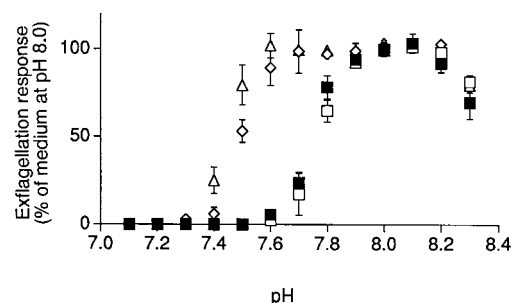


Figure 5 Xanthurenic acid extends the permissive pH range for exflagellation. Exflagellation responses over a pH range were compared in xanthurenic acid-free synthetic medium (RPMI 1640 with 40 mM HEPES and 20 mM sodium bicarbonate; filled squares), serum (pooled mouse sera, 40 mM HEPES; open squares), and serum supplemented with 3 μ M (diamonds) or 10 μ M (triangles) xanthurenic acid. Error bars show s.e.m. of 3 experiments.

found in mammalian serum at low (0.7 μ M) concentrations^{23,24}. We therefore compared exflagellation in sera from uninfected mice (Fig. 5, open squares) with that in synthetic medium without xanthurenic acid (Fig. 5, filled squares) *in vitro*. Both solutions supported exflagellation only over a narrow pH range (optimum pH 7.8–8.2) and failed to induce gametogenesis at normal blood pH (pH 7.3–7.4) or at the pH of the mosquito bloodmeal (pH 7.5–7.6; refs. 5, 6 and O.B., unpublished data). The addition of only 3 μ M xanthurenic acid to the serum greatly extended the permissive pH range (Fig. 5, diamonds). These data show that, in the mosquito bloodmeal, increased levels of xanthurenic acid could act together with the small rise in pH of 0.2–0.3 and the obligatory temperature drop following the ingestion of blood.

The putative role of xanthurenic acid as an inducer of malaria development thus depends critically on it occurring at elevated levels in the bloodmeal. We are seeking to develop a way of determining xanthurenic acid concentration in the bloodmeal by developing a stable isotope quantification method. GAF activity has been described in gut homogenate and in the anal secretions produced when mosquitoes are fed on saline containing phagostimulants^{7,12}, and this has been taken as indirect evidence for its presence in the gut lumen^{7,12}. However, because xanthurenic acid is excreted through the Malpighian tubules in *Drosophila*¹⁹, such interpretations must now be treated with extreme caution.

The identification of xanthurenic acid as a gametocyte-activating factor makes it possible to study the regulation of malaria gametogenesis at the cellular level. This may allow both the identification of the parasite receptors for xanthurenic acid and the further elucidation of secondary messenger pathways responsible for regulating gametogenesis. Gametogenesis is the first obligatory step in the mosquito transmission of these parasites. Therefore, these data could also lead to the development of new methods of interrupting malaria transmission by the use of chemical inhibitors or antagonists of xanthurenic acid activity, or by the selection of new mosquito genotypes which would be refractory to infection. □

Methods

Reverse-phase HPLC and HVPE. Extracts from 100–150 pupae were dissolved in 5% acetic acid and purified on a Kontron HPLC system using an Ultrasphere (250 $\times$ 4.6 mm) reverse-phase column by gradient elution²⁵ using solvent A (5% acetic acid) and increasing concentrations of solvent B (*n*-propanol) 0–20% for 10 min, 20–25% for 20 min and 25–100% for 5 min. High-voltage paper electrophoresis was carried out on Whatman no. 1 paper with either pH 6.5 pyridine/acetic acid or pH 2.1 formic acid/acetic acid buffers, at 3 kV for 30 min followed by drying and elution of strips using 5% acetic acid.

Mass spectrometry and derivative formation. Electrospray mass spectra of fractions (5–20%) of HPLC- or HVPE-purified biologically active material

were recorded in both positive- and negative-ion MS and MS/MS modes on a novel geometry Q-TOF instrument as described^{9–11}. Accurate mass measurements were determined in the negative-ion mode by fast atom bombardment on a ZAB2SE mass spectrometer tuned to 10,000 resolving power. Derivative-forming reactions on GAF with mass spectrometric monitoring in both MS and MS/MS modes to determine the number and nature of functional groups in the molecule were carried out as described^{13–15}.

Bioassay. Gametocytes of *Plasmodium berghei* ANKA strain clone 2.34 were produced in mice and used on days 3–5 post-inoculation. To test for GAF activity, 2.5 μ l of fresh tailblood was mixed immediately with 7.5 μ l of the test solution; 15–25 min later, exflagellating microgametocytes in the mixture were quantified as described⁸. HPLC fractions and synthetic compounds were tested in RPMI 1640 and 20 mM HEPES. A modification of the assay was used for the experiments shown in Fig. 4: 2 μ l of blood containing mature gametocytes was incubated for 15 min in 0.1 ml of test solution. Solutions were supplemented with additional HEPES (to 40 mM), and exflagellation was quantified in a haemocytometer and expressed as a percentage of exflagellation induced by a reference solution. Under these conditions the desired pH in the bicarbonate-buffered solutions was stable over the time of the experiment.

Received 11 November 1997; accepted 13 January 1998.

- Sinden, R. E. Sexual development of malarial parasites. *Adv. Parasitol.* **22**, 153–216 (1983).
- Sinden, R. E., Butcher, G., Billker, O. & Fleck, S. Regulation of infectivity of *Plasmodium* to the mosquito vector. *Adv. Parasitol.* **38**, 53–117 (1996).
- Carter, R. & Graves, P. M. in *Malaria—Principles and Practice of Malariology* (eds Wernsdorfer, W. H. & McGregor, I.) 273–305 (Churchill Livingstone, Edinburgh, 1988).
- Nijhout, M. M. & Carter, R. Gamete development in malaria parasites: bicarbonate-dependent stimulation by pH *in vitro*. *Parasitology* **76**, 39–53 (1978).
- Micks, D. W., de Caires, P. F. & Franco, L. B. The relationship of exflagellation in avian plasmodia to pH and immunity in the mosquito. *Am. J. Hyg.* **48**, 182–190 (1948).
- Bishop, A. & McConnachie, E. W. A study of the factors affecting the emergence of the gametocytes of *Plasmodium gallinaceum* from the erythrocytes and the exflagellation of the male gametocytes. *Parasitology* **46**, 192–215 (1956).
- Nijhout, M. M. *Plasmodium gallinaceum*: Exflagellation stimulated by a mosquito factor. *Exp. Parasitol.* **48**, 75–80 (1979).
- Billker, O., Shaw, M. K., Margos, G. & Sinden, R. E. The roles of temperature, pH and mosquito factors as triggers of male and female gametogenesis of *Plasmodium berghei* *in vitro*. *Parasitology* **14**, 1–7 (1997).
- Morris, H. R. *et al.* High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **10**, 889–896 (1996).
- Morris, H. R., Paxton, T., Panico, M., McDowell, R. & Dell, A. A novel geometry mass spectrometer, the Q-TOF, for low-femtomole/attomole-range biopolymer sequencing. *J. Protein Chem.* **16**, 469–479 (1997).
- Morris, A., Paxton, T., Panico, M., McDowell, R. & Dell, A. in *Mass Spectrometry of Biological Materials* Vol. 2 (eds Larsen, B. & McEwen, C.) 53–80 (Marcel Dekker, New York, 1998).
- Garcia, G. E., Wirtz, R. A. & Rosenberg, R. Isolation of a substance from the mosquito that activates *Plasmodium* fertilization. *Mol. Biochem. Parasitol.* **88**, 127–135 (1997).
- Morris, H. R., Taylor, G. W., Piper, P. J. & Tipples, J. R. Structure of slow-reacting substance of anaphylaxis from guinea-pig lung. *Nature* **285**, 104–106 (1980).
- Morris, H. R., Taylor, G. W., Masento, M. S., Jermyn, K. A. & Kay, R. R. Chemical structure of the morphogen differentiation inducing factor from *Dictyostelium discoideum*. *Nature* **328**, 811–814 (1987).
- Hunt, E. & Morris, H. R. Collagen cross-links: a mass spectrometric and ¹H- and ¹³C-nuclear-magnetic-resonance study. *Biochem. J.* **135**, 833–843 (1973).
- Battersby, A. R., Jones, K., McDonald, E., Robinson, J. A. & Morris, H. R. The structures and chemistry of isobacteriochlorins from *Desulphovibrio gigas*. *Tetrahedron Lett.* **25**, 2213–2216 (1977).
- Kayser, H. in *Comprehensive Insect Physiology, Biochemistry and Pharmacology* Vol. 10 (eds Kerkut, G. A. & Gilbert, L. I.) 368–416 (Pergamon, Oxford, 1985).
- Beard, C. B., Benedict, M. Q., Primus, J. P., Finnerty, V. & Collins, F. H. Eye pigments in wild-type and eye-color mutant strains of the African malaria vector *Anopheles gambiae*. *J. Hered.* **86**, 375–380 (1995).
- Wessing, A. & Eichelberg, D. Die fluoreszierenden Stoffe aus den Malpighischen Gefäßen verschiedener Augenfarbenmutanten von *Drosophila melanogaster*. *Z. Naturforsch.* **23**, 376–386 (1968).
- Phillips, J. P. & Forrest, H. S. in *The Genetics and Biology of Drosophila* Vol. 2 (eds Ashburner, M. & Wright, T. R. F.) 542–624 (Academic, London, 1980).
- Sullivan, D. T. & Sullivan, M. C. Transport defects as the physiological basis for eye color mutants of *Drosophila melanogaster*. *Biochem. Genet.* **13**, 603–613 (1975).
- Sullivan, D. T., Bell, L. A., Paton, D. R. & Sullivan, M. C. Purine transport by malpighian tubules of pteridine-deficient eye color mutants of *Drosophila melanogaster*. *Biochem. Genet.* **17**, 565–573 (1979).
- Truscott, J. W. R. & Elderfield, J. Relationship between serum tryptophan and tryptophan metabolite levels after tryptophan ingestion in normal subjects and age-related cataract patients. *Clin. Sci.* **89**, 591–599 (1995).
- Williams, S. A., Monti, J. A., Boots, L. R. & Cornwell, P. E. Quantitation of xanthurenic acid in rabbit serum using high-performance liquid chromatography. *Am. J. Clin. Nutr.* **40**, 159–167 (1984).
- Morris, H. R., Etienne, A. R., Dell, A. & Albuquerque, R. A rapid and specific method for the high resolution purification and characterization of neuropeptides. *J. Neurochem.* **34**, 574–582 (1980).
- Cochran, D. G. in *Comprehensive Insect Physiology, Biochemistry and Pharmacology* Vol. 4 (eds Kerkut, G. A. & Gilbert, L. I.) 467–506 (Pergamon, Oxford, 1985).

Acknowledgements. We thank A. Reason and J. Redfern (M-Scan) for MALDI Voyager Elite studies. This work was supported by the UNDP/World Bank/WHO special programme for research and training in tropical diseases (R.E.S.), the German Academic Exchange Service (HSP/III grant to O.B.) and the Wellcome Trust (H.R.M., A.D.).

Correspondence and requests for materials should be addressed to R.E.S. or H.R.M.

Genetic basis and molecular mechanism for idiopathic ventricular fibrillation

Qiuyun Chen^{†‡}, Glenn E. Kirsch^{†‡}, Danmei Zhang^{*}, Ramon Brugada[§], Josep Brugada^{||}, Pedro Brugada[¶], Domenico Potenza[#], Angel Moya^{*}, Martin Borggreffe^{**}, Günter Breithardt^{**}, Rocio Ortiz-Lopez^{*}, Zhiqing Wang^{†‡}, Charles Antzelevitch^{‡‡}, Richard E. O'Brien^{*}, Eric Schulze-Bahr^{**}, Mark T. Keating^{§§}, Jeffrey A. Towbin^{*†|||} & Qing Wang^{*}

^{*} Department of Pediatrics (Cardiology), [§] Medicine (Cardiology),

^{†‡} Cardiovascular Sciences, ^{|||} Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA

[‡] The Rammelkamp Center for Research, MetroHealth Campus, Case Western Reserve University, Cleveland, Ohio 44109, USA

^{||} Arrhythmia Unit, Cardiovascular Institute, Hospital Clinic, University of Barcelona, 08036 Barcelona, Spain

[¶] the Cardiovascular Center, OLV Hospital, 9300 Aalst, Belgium

[#] Cardiology Department, IRCCS Casa Sollievo della Sofferenza, S. Giovanni Rotondo, Italy

^{*} Department of Cardiology, Hospital Vall d'Hebron, 08020 Barcelona, Spain

^{**} Department of Cardiology and Angiology, Institute of Arteriosclerosis Research, Hospital of the University of Munster, D48129 Munster, Germany

^{‡‡} Masonic Medical Research Laboratory, Utica, New York 13504, USA

^{§§} Howard Hughes Medical Institute, Department of Human Genetics and Medicine, University of Utah, Salt Lake City, Utah 84112, USA

[†] These authors contributed equally to this work.

Ventricular fibrillation causes more than 300,000 sudden deaths each year in the USA alone^{1,2}. In approximately 5–12% of these cases, there are no demonstrable cardiac or non-cardiac causes to account for the episode, which is therefore classified as idiopathic ventricular fibrillation (IVF)^{3–6}. A distinct group of IVF patients has been found to present with a characteristic electrocardiographic pattern^{7–15}. Because of the small size of most pedigrees and the high incidence of sudden death, however, molecular genetic studies of IVF have not yet been done. Because IVF causes cardiac rhythm disturbance, we investigated whether malfunction of ion channels could cause the disorder by studying mutations in the cardiac sodium channel gene *SCN5A*. We have now identified a missense mutation, a splice-donor mutation, and a frameshift mutation in the coding region of *SCN5A* in three IVF families. We show that sodium channels with the missense mutation recover from inactivation more rapidly than normal and that the frameshift mutation causes the sodium channel to be non-functional. Our results indicate that mutations in cardiac ion-channel genes contribute to the risk of developing IVF.

Ventricular fibrillation is the most common cause of sudden cardiac death. When it occurs, there is no cardiac output, peripheral pulses and blood pressure are absent, and the patient loses consciousness. Death is imminent unless the arrhythmia is immediately controlled. IVF patients account for 3% of survivors of out-of-hospital cardiac arrest¹⁶. For patients who survive an episode of IVF, the rate of recurrence can be as high as 25–30% (refs 7, 17). A distinct electrocardiogram (ECG) feature, with right bundle branch block (RBBB) and elevation in ST segment in leads V1 to V3, has been described for some patients with sudden and aborted cardiac death due to IVF^{7–15}. Although the exact prevalence of IVF associated with RBBB and ST segment elevation is not known, clinical experience in Belgium and Spain suggests that it accounts for 40–60% of all IVF cases (data not shown).

We studied six small families and two sporadic patients with IVF by using single-strand conformation polymorphism (SSCP) and

DNA sequence analyses to identify mutations in known ion channel genes, including the cardiac sodium channel gene *SCN5A*. Two aberrant SSCP conformers were identified in all affected members in family K005, one in exon 21 of *SCN5A* (data not shown) and the other in exon 28 (Fig. 1a, b)¹⁸. Neither of the SSCP anomalies was seen in unaffected individuals (Fig. 1b) or in DNA samples from more than 150 control individuals (data not shown). DNA sequence analysis revealed two C-to-T base substitutions, one in exon 21 and the other in exon 28 (Fig. 1c): these mutations lead to substitution of an arginine by a tryptophan at codon 1,232 (represented as R1232W; data not shown), which lies in the extracellular loop between transmembrane segments S1 and S2 of domain III, and to substitution of a highly conserved threonine by a methionine at codon 1,620 (T1620M; Fig. 1c) in the extracellular loop between S3 and S4 of domain IV. Studies with sodium-channel-specific toxins indicated that the extracellular loop between DIVS3 and DIVS4 is important for coupling channel activation to fast inactivation^{19–21}.

Additional *SCN5A* mutations were found in two IVF families, K007 and K2823. Aberrant SSCP conformers were identified in affected individuals of K007 (Fig. 2a) and K2823 (Fig. 2b), but not in DNA samples from more than 150 control individuals (data not shown). DNA sequencing revealed that the abnormal SSCP band identified in K007 contained an insertion of two nucleotides, AA, which disrupted the splice-donor sequence of intron 7 (refs 18, 22) (Fig. 2a). The abnormal band identified in K2823 contained a deletion of a single nucleotide (A) at codon 1,397 (Fig. 2b). This deletion results in an in-frame stop codon that eliminates DIII/S6, DIV/S1–S6, and the carboxy-terminal portion of the cardiac sodium channel (Fig. 2b).

The potential contribution of R1232W and T1620M mutations to the mechanism of IVF was determined by heterologous expression

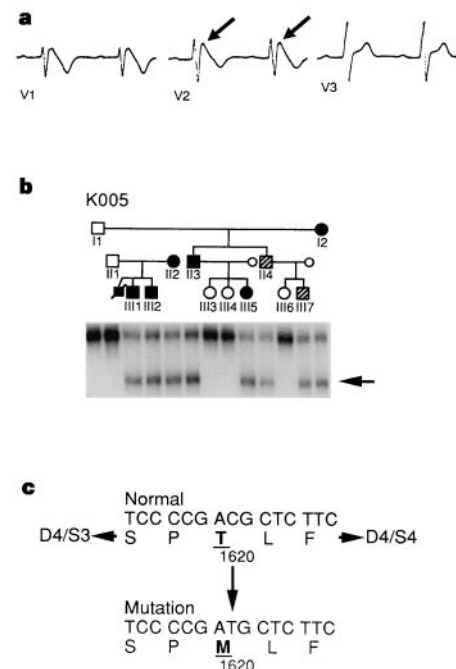


Figure 1 *SCN5A* missense mutation co-segregating with IVF in family K005. **a**, Electrocardiogram of an affected individual (III1). Note the ECG pattern of RBBB and ST segment elevation (diagonal arrows) in leads V1–V3. **b**, Pedigree structure and SSCP analysis with primers amplifying exon 28 of *SCN5A*. Affected individuals are filled circles (females) and filled squares (males). Unaffected individuals are empty symbols, and individuals without clinical data are shown as hatched. The individual who suffered sudden death from IVF is slashed. **c**, DNA and amino-acid sequences of the *SCN5A* missense mutation associated with IVF in K005. DNA sequence analysis revealed a C-to-T substitution which causes the T1620M mutation in the extracellular loop between DIVS3 and DIVS4.

in *Xenopus* oocytes. Current–voltage families were recorded from wild-type (WT) and mutated (R/W + T/M, containing both amino-acid substitutions R1232W and T1620M) channels. Peak conductance–voltage plots that represent steady-state activation (Fig. 3a) were indistinguishable between the two types of channel. The voltage dependence of steady-state inactivation (Fig. 3a), however, was shifted nearly 10 mV towards more positive potentials in R/W + T/M channels compared with WT channels, with no change in the slope of the curve. When fitted to a Boltzmann distribution, the midpoint potentials were -72.6 and -63.4 mV respectively in WT and R/W + T/M channels. We observed no other differences in steady-state gating properties of the two channels. In particular, unlike *SCN5A* mutations associated with long-QT syndrome^{23–26} (Fig. 4), no persistent, inactivation-resistant currents were observed in R/W + T/M channels (data not shown). These data confirm that IVF with RBBB and ST-segment elevation is a defect distinct from long-QT syndrome, which is a cardiac repolarization disorder characterized by a prolonged Q–T interval on ECG and a specific ventricular tachycardia, *torsade de pointes*.

Shifts in the voltage dependence of inactivation, unaccompanied by corresponding shifts in activation, suggested a change in the kinetics of inactivation. Therefore, we measured the time course of recovery from inactivation at negative potentials (Fig. 3b). At a recovery potential (V_r) of -80 mV, R/W + T/M channels showed a significantly faster recovery than wild-type channels. The time constant of the recovery process was markedly accelerated by more negative recovery potentials, and the difference between mutant and normal channels disappeared at potentials more negative than -110 mV. In channels that were mutated at either R1232 or T1620, the R1232W mutant behaved most like normal channels whereas the T1620M mutant closely followed the kinetic pattern of the double (R/W + T/M) mutant (Fig. 3b). This indicates that T1620M is the mutation that is probably responsible for the IVF phenotype and that R1232W could be a rare polymorphism. In summary, biophysical analysis of the two missense mutations in *SCN5A* showed a shift in the voltage dependence of steady-state inactivation towards more positive potentials associated with a 25–30% acceleration in recovery time from inactivation at potentials near -80 mV.

Genetic analysis demonstrated that all affected individuals in K005 carried two substitutions (T1620M and R1232W) on one chromosome and none on the other chromosome, suggesting that IVF in this family is inherited as an autosomal dominant trait. The presence of both normal and mutated sodium channels in the same tissue would promote heterogeneity of the refractory period, a well established mechanism in arrhythmogenesis, and therefore may be the underlying molecular defect that causes re-entrant arrhythmia in the IVF family K005.

The functional consequences of the splicing mutation have not been established. In view of the location of the mutation, within the intracellular loop between DIS2 and DIS3 (Fig. 4), it may reduce the number of functional sodium channels in IVF patients. The one-base-pair (1-bp) deletion in K2823 results in an intragenic stop codon which generates a truncated sodium channel with transmembrane domains DI, DII and part of DIII (Fig. 4). When tested in *Xenopus* oocytes, mutant RNA failed to express sodium currents (data not shown). Similar findings have been reported in rat brain sodium channels²⁷. These results indicate that all four transmembrane domains are required for formation of a functional sodium channel.

Inhibition of the sodium channel current causes heterogeneous loss of the action potential dome in the right ventricular epicardium, leading to a marked dispersion of repolarization and refractoriness, an ideal substrate for the development of re-entrant arrhythmia^{28,29}. Phase 2 re-entry produced by the same substrate is thought to provide the premature beat that initiates ventricular tachycardia and fibrillation. Both the splicing mutation and the

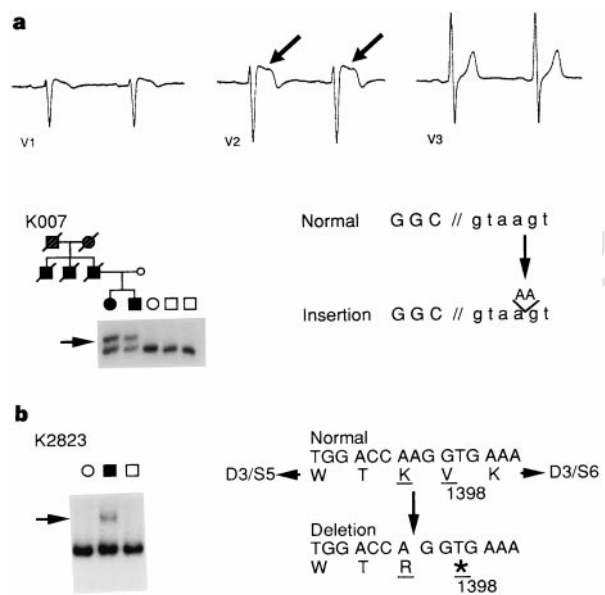


Figure 2 *SCN5A* splicing mutation in K007 and intragenic deletion in K2823. **a**, *SCN5A* splicing mutation in K007 and electrocardiogram obtained from the male affected individual (proband) in K007. Pedigree structure and the abnormal SSCP conformers (indicated by an arrow) are shown. DNA sequence analysis of the normal and abnormal conformers in K007 revealed a two-base-pair (AA) insertion. This insertion occurs in the splicing donor site of intron 7 of *SCN5A*. The symbol // indicates the exon–intron boundary. **b**, A deletion of a single nucleotide A was identified in the abnormal SSCP band of K2823. The deletion causes a frameshift and leads to the premature termination of the cardiac sodium channel starting in the pore region of transmembrane domain III. The electrocardiogram for the proband in K2823 is not shown because the patient presented mild J point depression anteriorly with a rapidly upsloping ST segment during a peak exercise, but not at the baseline.

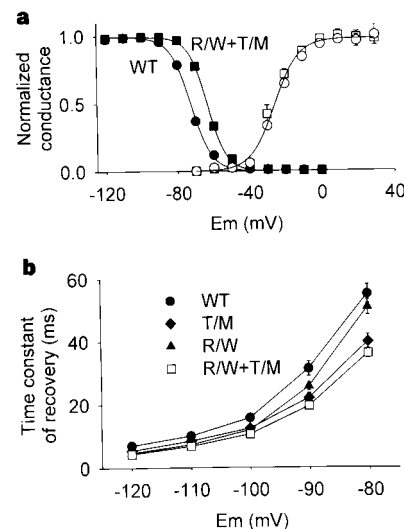


Figure 3 Voltage-dependence of activation and inactivation, and time course of recovery from inactivation. **a**, Steady-state voltage-dependence of activation and inactivation in normal and mutated sodium channels. Open symbols: activation in wild-type (WT, open circles; $n = 7$) and mutant (R/W + T/M, open squares; $n = 7$) channels fit to a Boltzmann distribution with midpoint -25.4 mV and slope factor 6.5 mV. Filled symbols: inactivation in WT (filled circles; $n = 7$) and mutant channels (filled squares; $n = 7$; error bars smaller than symbols) fit to Boltzmann distributions with midpoint potentials of -72.6 and -63.4 mV (slope factors of 5.5 and 5.2 mV), respectively in WT and R/W + T/M channels. **b**, Time course of recovery from inactivation. Average time constants versus recovery potential (V_r) for WT (filled circles; $n = 9$) and mutant channels: T/M (filled diamonds; $n = 6$), R/W (filled triangles; $n = 8$) and R/W + T/M (open squares; $n = 5$).

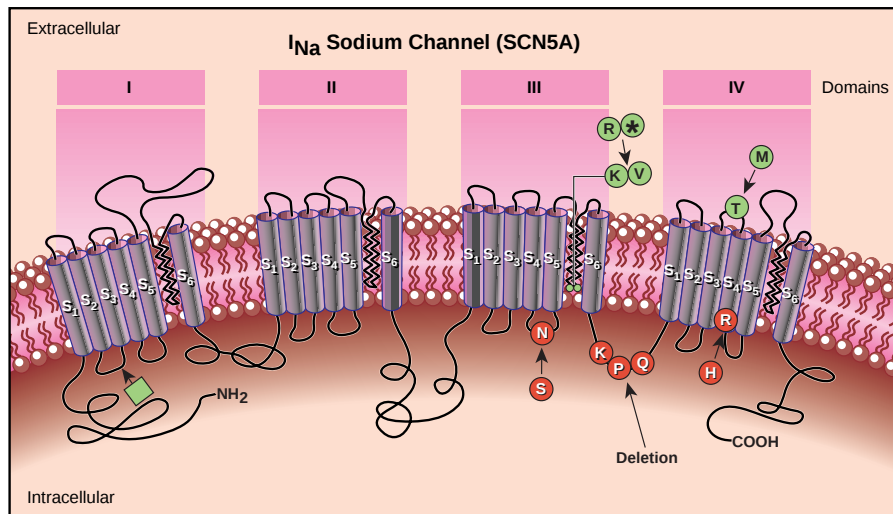


Figure 4 The predicted secondary structure of the cardiac sodium channel and locations of mutations causing IVF and chromosome-3-linked long-QT syndrome (LQT). The channel consists of four putative transmembrane domains (DI–DIV),

one-nucleotide deletion result in a reduction in the number of functional sodium channels, which should promote the development of re-entrant arrhythmias.

This work shows that IVF with RBBB and ST segment elevation is a distinct syndrome and that mutations in the cardiac sodium channel gene *SCN5A* are associated with this disorder. We identified two types of *SCN5A* mutation in IVF patients (Fig. 4): some patients carry a missense mutation that leads to altered *SCN5A* function, and others seem to carry loss-of-function mutations. It is unclear how two types of mutation with different functional effects on sodium channels can lead to a similar IVF syndrome: both types cause ventricular fibrillation in individuals with a structurally normal heart, presumably as a result of re-entrant mechanisms; however, ECGs showing ST segment elevation and RBBB are not homogeneous—the ST segment elevation may exist only as J-point elevation³⁰ (as in family K005) or as complete ST segment elevation (as in family K007). Whether these differences in the ECG pattern are due to specific mutations or other modifying factors is unknown, and careful phenotype–genotype correlation of more IVF families is needed to clarify this further. Genetic testing and a mechanistic understanding of IVF may lead to rational therapies for cardiac arrhythmias. □

Methods

Identification and phenotyping of kindreds with IVF. IVF kindreds were ascertained from medical clinics in North America, Germany, Spain, Italy and Belgium. Phenotypic criteria were identical to those used previously^{7,8}. Diagnosis was defined by previous history of sudden death in probands or in a family member, no demonstrable structural heart disease or long-QT syndrome by invasive and non-invasive procedures, and an ECG pattern of RBBB and ST segment elevation in leads V1–V3 at baseline or after intravenous Ajmaline, as described previously^{7–10}. The proband in K005 had a twin brother who died at the age of 23 years from IVF. In K007, the three brothers died of IVF at ages of 43, 46 and 35 years, respectively. The father of the proband had two syncopal episodes at ages 20 and 34 years before his fatal episode one year later.

The proband in K2823 had four syncopal episodes, which was diagnosed as IVT. The proband's father died suddenly from cardiac arrest at the age of 48 years, and a cousin died from a heart attack at the age of 27 years. IVT experienced by the proband was inducible by electrophysiological studies, and degenerated into ventricular fibrillation which required direct current cardioversion. The proband did not present the typical ECG pattern of RBBB and ST segment elevation at baseline.

SSCP and DNA sequence analyses. SSCP and DNA analyses were done as before^{23,24} using PCR primers described previously¹⁸.

with each domain containing six transmembrane segments (S1–S6). IVF mutations are shown in green and LQT-associated mutations are in red.

Electrophysiological characterization of human cardiac sodium channels. Site-directed mutagenesis, *in vitro* transcription, *Xenopus* oocyte injection, electrophysiology, and data analysis were carried out as described²⁶. Activation was estimated from current–voltage families evoked by step pulses of –50 to +50 mV (10-mV increment) from a holding potential at –100 mV. Peak test pulse currents were converted to conductances ($G = I/(V - V_{rev})$, where G is the conductance, I is the peak current, V is the test pulse and V_{rev} is the measured reversal potential). Conductance was normalized against the maximum observed in each cell and the mean values at each test potential were plotted. Inactivation was determined using a standard two-pulse protocol in which a 500-ms conditioning potential to varying levels (–120 to 0 mV) was followed by a test step to –10 mV (20 ms) to assess the extent of inactivation produced during the conditioning step. Test pulse currents were normalized to the maximum observed current (at the most negative conditioning potentials) and mean values $\pm$ s.e.m. were plotted as a function of conditioning pulse amplitude. The time course of recovery from inactivation was measured using a three-pulse protocol in which a 500-ms conditioning pulse (V_c) to –10 mV, allowing complete inactivation, was followed by a variable recovery interval (V_r). Peak current evoked by test pulse (V_t ; –10 mV, 20 ms) was used to assay the amount of recovery. The envelope of superimposed test pulse currents was plotted versus the duration of the recovery interval and fitted to single exponential decay functions. The holding potential (V_h) was –120 mV.

Received 29 September; accepted 23 December 1997.

1. Kannel, W. B., Cupples, A. & D'Agostino, R. B. Sudden death risk in overt coronary heart disease: the Framingham study. *Am. Heart J.* **113**, 799–804 (1987).
2. Willich, S. et al. Circadian variation in the incidence of sudden cardiac death in the Framingham heart study population. *Am. J. Cardiol.* **60**, 801–806 (1987).
3. Trappe, H. J. et al. Prognosis of patients with ventricular tachycardia and ventricular fibrillation: role of the underlying etiology. *J. Am. Coll. Cardiol.* **12**, 166–174 (1988).
4. Viskin, S. & Belhassen, B. Idiopathic ventricular fibrillation. *Am. Heart J.* **120**, 661–671 (1990).
5. Wichter, T., Breithardt, G. & Borggrefe, M. In *Cardiac Arrhythmia: Mechanism, Diagnosis, and Management* (eds Podrid, P. J. & Kowey, P. R.) 1219–1238 (Williams & Wilkins, Maryland, 1995).
6. Martini, B. et al. Ventricular fibrillation without apparent heart disease: description of six cases. *Am. Heart J.* **118**, 1203–1209 (1989).
7. Brugada, J., Brugada, R. & Brugada, P. Right bundle branch block, ST segment elevation in leads V1–V3: a marker for sudden death in patients without demonstrable structural heart disease. *Circulation* **96**, 1151 (1997).
8. Brugada, P. & Brugada, J. Right bundle branch block, persistent ST segment elevation and sudden cardiac death: a distinct clinical and electrocardiographic syndrome. A multicenter report. *J. Am. Coll. Cardiol.* **20**, 1391–1396 (1992).
9. Brugada, P. & Brugada, J. Further characterization of the syndrome of right bundle branch block, ST segment elevation, and sudden cardiac death. *J. Cardiovasc. Electrophysiol.* **8**, 325–331 (1997).
10. Brugada, J., Brugada, P. & Brugada, R. Ajmaline unmasks right bundle branch block-like and ST segment elevation in V1–V3 in patients with idiopathic ventricular fibrillation. *PACE* **19**, 599 (1996).
11. Miyayama, H., Sakurai, M. & Odaka, H. Two cases of idiopathic ventricular fibrillation with interesting electrocardiographic findings. *Kokyo to Junkan* **41**, 287–291 (1993).
12. Aizawa, Y. et al. Idiopathic ventricular fibrillation and bradycardia-dependent intraventricular block. *Am. Heart J.* **126**, 1473–1474 (1993).
13. Sumiyoshi, M. et al. A case of idiopathic ventricular fibrillation with incomplete right bundle branch block and persistent ST segment elevation. *Jpn Heart J.* **34**, 661–666 (1993).

14. Bjerregaard, P., Gussak, I., Kotar, S. L., Gessler, J. E. & Janosik, D. Recurrent syncope in a patient with prominent J wave. *Am. Heart J.* **127**, 1426–1430 (1994).
15. Miyazaki, T. et al. Autonomic and antiarrhythmic drug modulation of ST segment elevation in patients with Brugada syndrome. *J. Am. Coll. Cardiol.* **27**, 1061–1070 (1996).
16. Tung, R. T., Shen, W., Hammill, S. C. & Gersh, E. J. Idiopathic ventricular fibrillation in out-of-hospital cardiac arrest survivors. *PACE* **17**, 1405–1411 (1994).
17. Cobb, L. A. in *Hurst's The Heart* (eds Schlant, R. C. & Alexander, R. W.) 8th edn. 947–957 (McGraw Hill, New York, 1994).
18. Wang, Q., Li, Z., Shen, J. & Keating, M. T. Genomic organization of the human *SCN5A* gene encoding the cardiac sodium channel. *Genomics* **34**, 9–16 (1996).
19. Rogers, J. C., Qu, Y., Tanada, T. N., Scheuer, T. & Catterall, W. A. Molecular determinants of high affinity binding of α -scorpion toxin and sea anemone toxin in the S3–S4 extracellular loop in domain IV of the Na channel α subunit. *J. Biol. Chem.* **271**, 15950–15962 (1996).
20. Hoffman, E. P., Lehmann-Horn, F. & Rudel, R. Overexcited or inactive: ion channels in muscle disease. *Cell* **80**, 681–686 (1995).
21. Yang, N. & Horn, R. Evidence for voltage-dependent S4 movement in sodium channels. *Neuron* **15**, 213–218 (1995).
22. Shapiro, M. B. & Senapathy, P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**, 7155–7174 (1987).
23. Wang, Q. et al. *SCN5A* mutations associated with an inherited cardiac arrhythmia, long QT syndrome. *Cell* **80**, 805–811 (1995).
24. Wang, Q. et al. Cardiac sodium channel mutations in patients with long QT syndrome, an inherited cardiac arrhythmia. *Hum. Mol. Genetics* **4**, 1603–1607 (1995).
25. Bennett, P. B., Yazawa, K., Makita, N. & George, A. L. Jr Molecular mechanism for an inherited cardiac arrhythmia. *Nature* **376**, 683–685 (1995).
26. Dumaine, R. et al. Multiple mechanisms of Na⁺ channel-linked long-QT syndrome. *Circulation Res.* **78**, 916–922 (1996).
27. Stuhmer, W. et al. Structural parts involved in activation and inactivation of the sodium channel. *Nature* **339**, 597–603 (1989).
28. Krishnan, S. C. & Antzelevitch, C. Sodium channel block produces opposite electrophysiological effects in canine ventricular epicardium and endocardium. *Circulation Res.* **69**, 277–291 (1991).
29. Krishnan, S. C. & Antzelevitch, C. Flecainide-induced arrhythmia in canine ventricular epicardium. Phase 2 reentry? *Circulation* **87**, 562–572 (1992).
30. Yan, G. X. & Antzelevitch, C. Cellular basis for the electrocardiographic J wave. *Circulation* **93**, 372–379 (1996).

Acknowledgements. We thank H. Hartmann for the wild-type *SCN5A* construct; M. Sanguinetti and P. Spector for help with electrophysiological analysis of the 1-bp deletion; and P. Szafranski, J. T. Bricker, M. Scheinman, A. L. Beudet, A. Bradley and X. Qu for help and advice. This work was supported by a Grant-In-Aid from the American Heart Association, by the AHA, Northeast Ohio Affiliate (G.E.K.), and the Deutsche Forschungsgemeinschaft (E.S.-B.), and by grants from the NIH and Bristol-Myers Squibb (M.T.K.), The Texas Children's Hospital Foundation Chair in Pediatric Cardiac Research and NIH grants (J.A.T.), the Carolien Weiss Law Grant for Research in Molecular Medicine (Q.W.), the Abercrombie Cardiology Fund of Texas Children's Hospital (Q.W.), and a Scientist Development Award from the American Heart Association (Q.W.).

Correspondence and requests for materials should be addressed to Q.W. (e-mail: qwang@bcm.tmc.edu).

Fas-mediated apoptosis and activation-induced T-cell proliferation are defective in mice lacking FADD/Mort1

Jianke Zhang, Dragana Cado, Ann Chen, Nisha H. Kabra & Astar Winoto

Department of Molecular and Cell Biology, Division of Immunology and Cancer Research Laboratory, University of California at Berkeley, 469 LSA, Berkeley, California 94720-3200, USA

Programmed cell death, or apoptosis, is important in homeostasis of the immune system: for example, non-functional or autoreactive lymphocytes are eliminated through apoptosis. One member of the tumour necrosis factor receptor (TNFR) family, Fas (also known as CD95 or Apo-1), can trigger cell death and is essential for lymphocyte homeostasis^{1,2}. FADD/Mort1 (refs 3–6) is a Fas-associated protein that is thought to mediate apoptosis by recruiting the protease caspase-8 (refs 7, 8). A dominant-negative mutant of FADD inhibits apoptosis initiated by Fas and other TNFR family members^{9–14}. Other proteins, notably Daxx, also bind Fas and presumably mediate a FADD-independent apoptotic pathway¹⁵. Here we investigate the role of FADD *in vivo* by generating FADD-deficient mice. As homozygous mice die *in utero*, we generated *FADD*^{−/−} embryonic stem cells and *FADD*^{−/−} chimaeras in a background devoid of the recombination activating gene *RAG-I*, which activates rearrangement of the immunoglobulin and T-cell receptor genes. We found that thymocyte subpopulations were apparently normal in newborn chimaeras.

Fas-induced apoptosis was completely blocked, indicating that there are no redundant Fas apoptotic pathways. As these mice age, their thymocytes decrease to an undetectable level, although peripheral T cells are present in all older *FADD*^{−/−} chimaeras. Unexpectedly, activation-induced proliferation is impaired in these *FADD*^{−/−} T cells, despite production of the cytokine interleukin (IL)-2. These results and the similarities between *FADD*^{−/−} mice and mice lacking the β -subunit of the IL-2 receptor suggest that there is an unexpected connection between cell proliferation and apoptosis.

To generate FADD-deficient mice we replaced the first exon of FADD with a neomycin-resistance (neo) gene (Fig. 1a). Heterozygous mutant embryonic stem (ES) cells were identified by Southern blot analysis (Fig. 1b), and *FADD*^{+/−} mice were produced from three *FADD*^{+/−} ES lines. Interbreeding of these mice, however, yielded only heterozygous and wild-type mice in a 2:1 ratio. Subsequent studies established that *FADD*^{−/−} embryos die at around day 9 of gestation (data not shown), indicating that FADD is essential for embryonic development.

Because Fas is necessary for homeostasis in the immune system, we sought to investigate the effect of FADD deletion in lymphoid organs. However, this can not be examined directly because *FADD*^{−/−} mice die *in utero*. We therefore made *FADD*^{−/−} ES cells and injected them into C57BL/6-*RAG-I*^{−/−} (B6-*RAG-I*^{−/−}) blastocysts to generate *FADD*^{−/−} chimaeras¹⁶. Homozygous *FADD*^{−/−} ES cells were obtained by subjecting *FADD*^{+/−} ES cells to a high level of G418 selection¹⁷ (Fig. 1b). Complete absence of FADD protein was confirmed by western blot analysis (Fig. 1c). Because these ES donor cells originate from 129/Sv strain (agouti colour) and host blastocysts originate from B6-*RAG-I*^{−/−} mice (black), chimaeras were readily identified by coat colour. By this means we generated 20 viable chimaeras. All mature lymphocytes in these *FADD*^{−/−} → *RAG-I*^{−/−} chimaeras (designated *FADD*^{−/−}) are derived from *FADD*^{−/−} ES cells because *RAG-I*^{−/−} mice are not capable of producing any B or T cells¹⁸. An antibody for the allelic Ly9.1 antigen provides another means of distinguishing the donor (129/Sv mice are Ly9.1⁺) from the host cells (B6 mice are Ly9.1[−]). Tissue Southern blot analysis confirmed the contribution of *FADD*^{−/−} ES cells in many organs and tissues (data not shown).

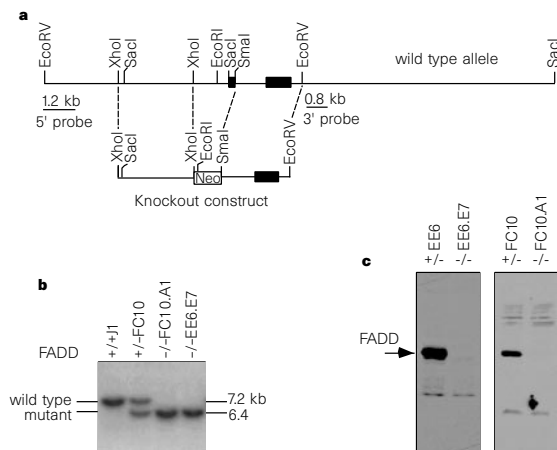


Figure 1 Targeted inactivation of the gene encoding FADD/Mort1 in ES cells and in mice. **a**, A restriction map of the murine *FADD* locus is shown (filled boxes denote exons)⁶. A schematic representation of the FADD targeting construct is shown underneath. **b**, Generation of FADD-deficient ES cell clones. Heterozygous ES cell clones were identified by Southern blots with 5' probe which detects a 7.2-kilobase (*EcoRI*–*EcoRV*) wild-type allele and a 6.4-kb mutant allele. Two *FADD*^{−/−} clones, FC10.A1 and EE6.E7, were obtained after further G418 selection¹⁷. They contain only the mutant alleles. J1 is the parental wild-type ES cell for FC10. **c**, The absence of FADD protein in *FADD*^{−/−} lines was confirmed by western blot analysis using anti-FADD antibodies⁶.

We analysed the cellular profile of thymi from *FADD*^{-/-} chimaeras. Thymic cellularity of newborn to 5-week-old *FADD*^{-/-} chimaeras ($10.8 \pm 2.3 \times 10^6$, $n = 7$) is not significantly greater than that of *RAG-1*^{-/-} mice ($9.3 \pm 2.9 \times 10^6$, $n = 5$). However, *RAG-1*^{-/-} chimaeras reconstituted with the wild-type ES cells have a similar number of thymic cells to wild-type mice. Thus *FADD*^{-/-} donor ES cells seem to be inefficient in reconstituting and/or maintaining thymic cellularity. T-cell development in *RAG-1*^{-/-} mice is blocked at the CD4⁺CD8⁺ stage¹⁸ (Fig. 2). In contrast, newborn and 4-day-old *FADD*^{-/-} chimaeras have thymocyte subsets resembling those of wild-type mice (Fig. 2). All CD4⁺CD8⁺ double-positive and single-positive cells are Ly9.1⁺, confirming their ES cell origin. As the chimaeras age, however, the population of *FADD*^{-/-} thymocytes greatly diminishes. By five weeks, few or no double-positive thymocytes remain (Fig. 2). Nevertheless, the mature Ly9.1⁺ T lymphocytes (CD3⁺) are present in peripheral organs (Fig. 3).

It is possible that the disappearance of thymocytes over time could result from an intrinsic survival defect of *FADD*^{-/-} cells. Alternatively, *FADD*^{-/-} cells might cause some physiological changes in the thymus, such as production of cytotoxic cytokines. This latter possibility is unlikely because chimaeras in wild-type background (*FADD*^{-/-} → B6) have a normal number of host Ly9.1⁺ thymocytes. *FADD*^{-/-} thymocytes seem to die faster *in vitro* than their wild-type counterparts (data not shown), suggesting that *FADD*^{-/-} thymocytes have an intrinsic defect.

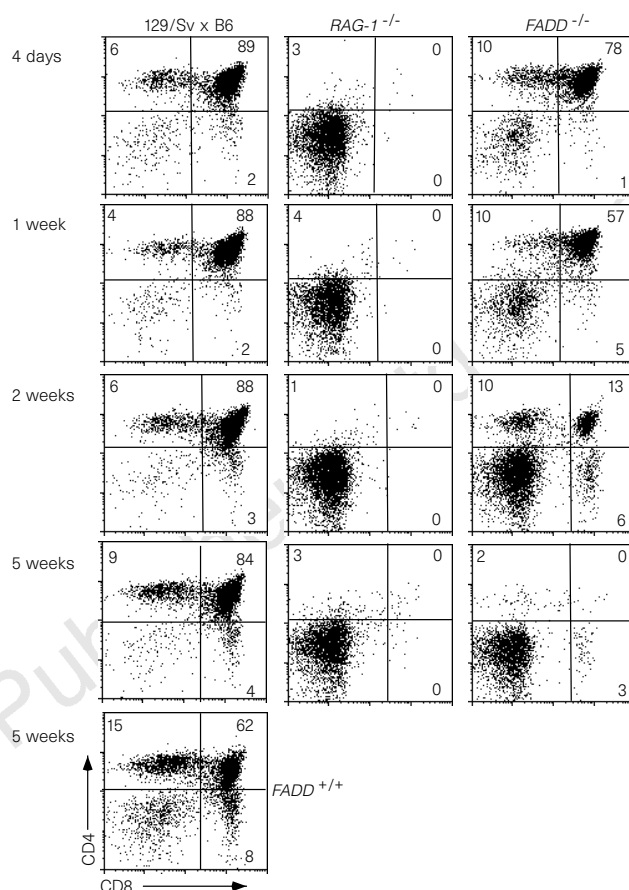


Figure 2 Flow cytometric analysis of thymocytes from wild-type 129/Sv x B6, *RAG-1*^{-/-}, *FADD*^{-/-} to *RAG-1*^{-/-} (*FADD*^{-/-}), and *FADD*^{+/+} to *RAG-1*^{-/-} (*FADD*^{+/+}) mice. Thymocytes from mice with ages indicated were analysed with anti-CD4 and anti-CD8 antibodies. The percentage of each population is indicated in each quadrant.

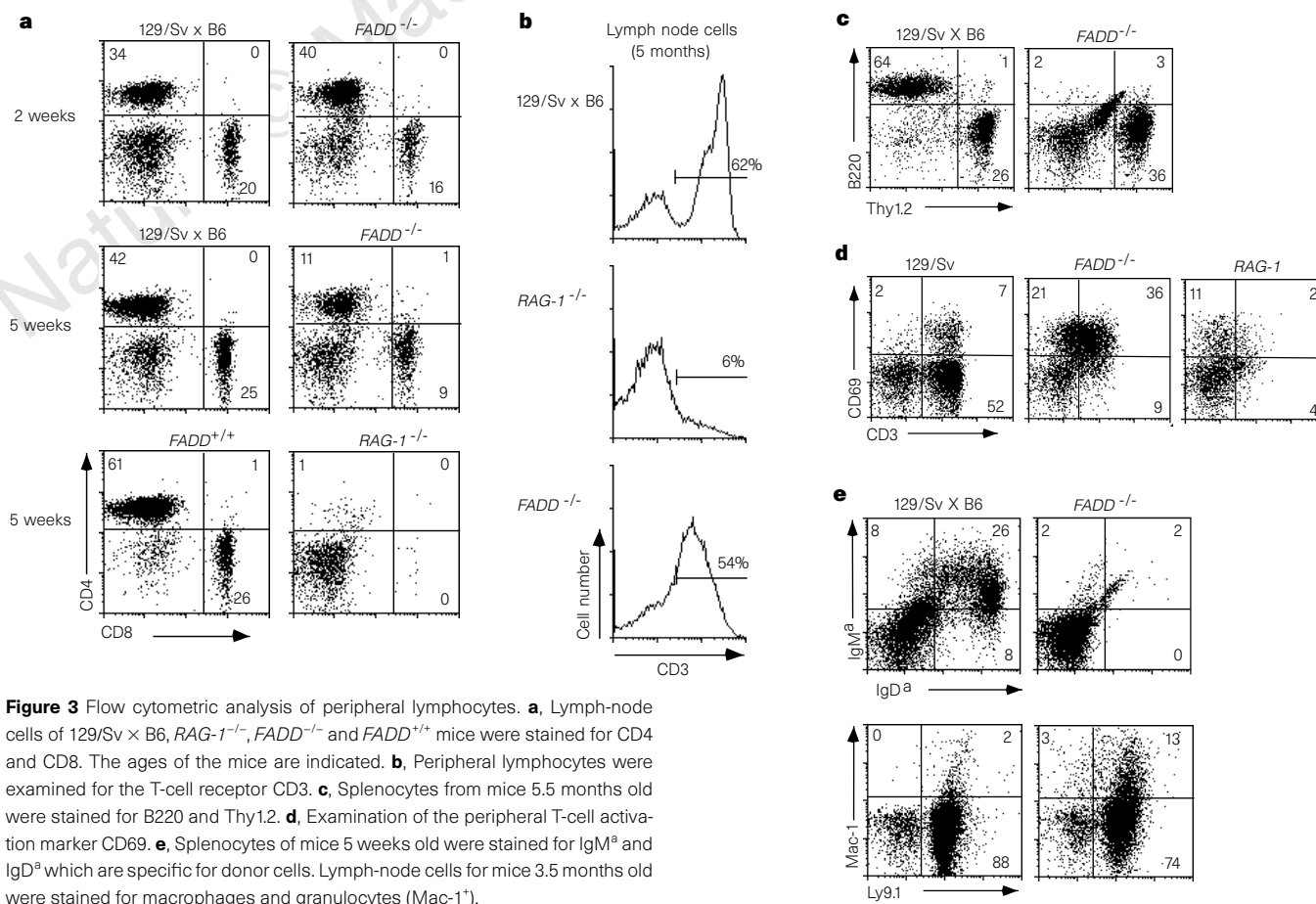
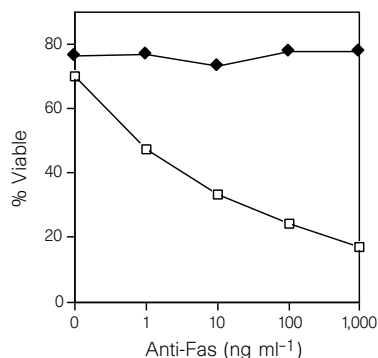


Figure 3 Flow cytometric analysis of peripheral lymphocytes. **a**, Lymph-node cells of 129/Sv x B6, *RAG-1*^{-/-}, *FADD*^{-/-} and *FADD*^{+/+} mice were stained for CD4 and CD8. The ages of the mice are indicated. **b**, Peripheral lymphocytes were examined for the T-cell receptor CD3. **c**, Splenocytes from mice 5.5 months old were stained for B220 and Thy1.2. **d**, Examination of the peripheral T-cell activation marker CD69. **e**, Splenocytes of mice 5 weeks old were stained for IgM^a and IgD^a which are specific for donor cells. Lymph-node cells for mice 3.5 months old were stained for macrophages and granulocytes (Mac-1^a).

In contrast to the thymus, *FADD*^{-/-} mice contain large numbers of T cells in peripheral lymphoid organs. The total number of *FADD*^{-/-} splenocytes ($23.4 \pm 9.6 \times 10^6$, $n = 8$) is reduced roughly twofold compared with that of the wild-type counterparts ($54.5 \pm 13.1 \times 10^6$, $n = 7$); *RAG*-1^{-/-} mice have fewer splenocytes ($8.73 \pm 3.1 \times 10^6$, $n = 6$). The lymph-node cell number of *FADD*^{-/-} mice is similar to that of wild-type mice (*FADD*^{-/-}, $28.3 \pm 8.6 \times 10^6$, $n = 6$; wild type, $36.4 \pm 3.7 \times 10^6$, $n = 5$; *RAG*-1^{-/-}, $4.7 \pm 3.1 \times 10^6$, $n = 4$). Single-positive (CD3⁺) T cells are present in both *FADD*^{-/-} lymph nodes (Fig. 3a, b) and spleens.



◀ **Figure 4** Fas-mediated cell death is blocked in the absence of FADD. Thymocytes from *FADD*^{+/+} (open squares) and *FADD*^{-/-} (filled diamonds) mice 4 days old were incubated with increasing concentrations of the anti-Fas antibody Jo2 in the presence of cycloheximide as described²². Viability was determined by propidium iodide exclusion. Values are averages of duplicates. The same analysis was performed on mice 1 and 2 weeks old with similar results.

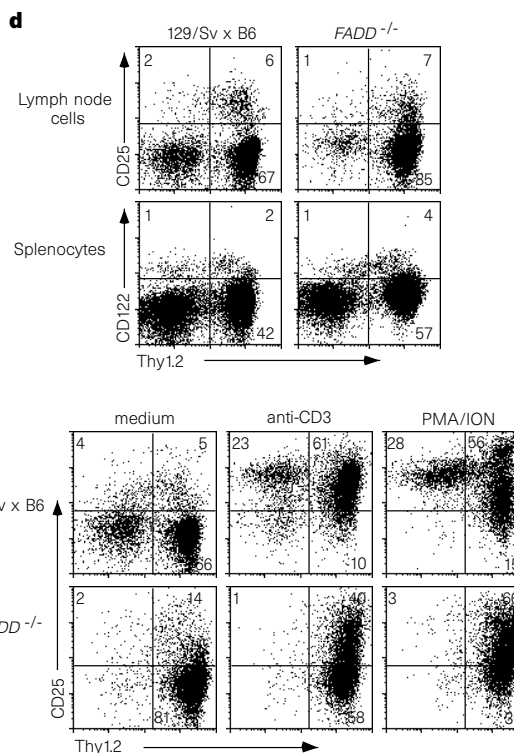
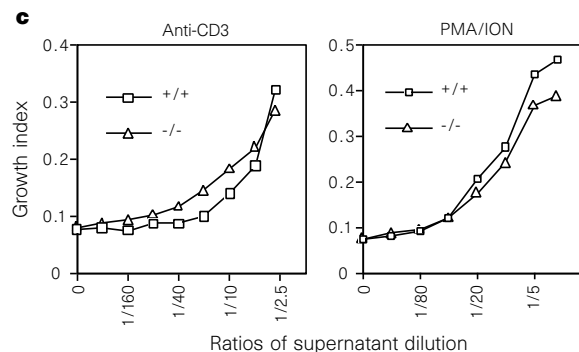
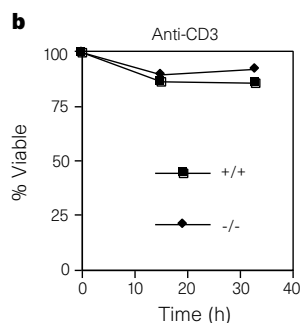
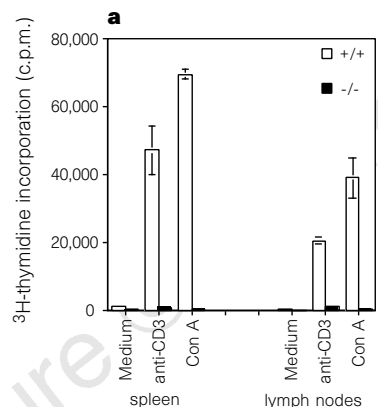


Figure 5 T-cell proliferation and death and IL-2 secretion assay. **a**, Lymphocytes from *FADD*^{+/+} (white bars) and *FADD*^{-/-} (black bars) mice 3.5 months old were treated with anti-CD3 antibody or concanavalin A. Proliferation was monitored by [³H]thymidine incorporation. The experiment was done in triplicate. Similar results were observed with mice two weeks old. **b**, Cell viability of anti-CD3 stimulated lymph-node T cells (Thy1.2⁺) was followed by staining with 7-amino-actinomycin D and flow cytometry as described³⁰. +/+, squares; -/-, diamonds. **c**, IL-2 production from stimulated splenocytes was determined by CTL2-2 survival and MTT assay. **d**, Unstimulated (top) or stimulated (bottom) lymph-node cells were stained for CD25 and splenocytes for CD122.

IgM^a and IgD^a, which are allele-specific for the 129/Sv strain, to distinguish donor from host cells. Strikingly, *FADD*^{-/-} chimaeras do not contain detectable levels of Ly9.1⁺B220⁺IgM^aIgD^aB cells in either bone marrow or spleen (Fig. 3c, e), which may indicate that deletion of *FADD* leads to a blockage of B-cell development. Using Ly9.1 as a donor cell marker, we examined other haematopoietic cells in *FADD*^{-/-} chimaeras. Macrophages and granulocytes (Ly9.1⁺Mac-1⁺) are present in *FADD*^{-/-} mice (Fig. 3e). The high proportion of Ly9.1⁺Mac-1⁺ cells in these *FADD*^{-/-} mice is probably due to the absence of B lymphocytes.

Proteins other than *FADD* have also been reported to bind to Fas^{15,19–21}. Expression of one of these Fas-interacting proteins, Daxx, potentiates Fas-induced apoptosis in an apparently *FADD*-independent manner, suggesting that there are parallel apoptotic pathways downstream of Fas¹⁵. To investigate this, we incubated *FADD*^{-/-} thymocytes with an agonist anti-Fas antibody²². Both thymic and peripheral T cells in *FADD*^{-/-} chimaeric mice express Fas at a high level, comparable to that of wild-type T cells (data not shown). Wild-type thymocytes are killed in a dose-dependent manner (Fig. 4). In contrast, thymocytes from *FADD*^{-/-} mice are completely resistant to the cytotoxic effect of anti-Fas antibody. Thus *FADD* is an essential mediator of Fas-induced apoptosis and there is unlikely to be a parallel apoptotic pathway downstream of Fas. Although thymocytes are less sensitive to TNF-mediated apoptosis (twofold effect)²³, our results also showed that *FADD*^{-/-} thymocytes are more resistant to this than their wild-type counterparts (data not shown).

The absence in *FADD*^{-/-} mice of lymphoproliferative disease and the T-cell population characteristic of Fas mutant mice may result from a defect in the activation and proliferation of T cells. To examine this possibility, peripheral lymphocytes of age-matched wild-type and *FADD*^{-/-} mice were stimulated with an anti-CD3 antibody or concanavalin A (Con A). Wild-type T cells proliferated extensively, whereas *FADD*^{-/-} peripheral T cells did not (Fig. 5a). The lack of proliferation is not due to the death of *FADD*^{-/-} peripheral T cells during stimulation, as no differences in cell viability were observed between wild-type and mutant cells (Fig. 5b).

To establish whether the reduced proliferation of *FADD*^{-/-} T cells results from the lack of IL-2 production or unresponsiveness to IL-2, we assayed for the secretion of IL-2. After stimulation, *FADD*^{-/-} cells produced wild-type levels of IL-2, as shown by bioassay and enzyme-linked immunosorbent assay (ELISA) (Fig. 5c and data not shown). Freshly isolated mutant T cells have normal levels of IL-2R α (CD25) and IL-2R β (CD122) (Fig. 5d, top). Activation of T cells leads to a dramatic induction of CD25 expression in both wild-type and mutant T cells (Fig. 5d, bottom). Thus *FADD*^{-/-} T cells fail to proliferate despite there being normal levels of IL-2R and IL-2 production.

We have shown that *in vivo* *FADD* is essential for Fas-induced apoptosis and that there is no parallel Fas apoptotic pathway. As mice with mutations in either Fas or TNF-R1 are viable, we were surprised to find that deletion of *FADD* leads to embryonic lethality. Because other apoptotic receptors can interact with *FADD*^{10–14}, their function may be necessary during embryogenesis. Alternatively, *FADD* may be involved in other signalling pathways required for development and proliferation, as *FADD* seems to be required for T-cell proliferation. Although these phenotypes could arise from abnormal T-cell development in the absence of *FADD*, we believe that they reflect direct roles of *FADD* *in vivo*. Several characteristics of *FADD*-deficient chimaeras are strikingly similar to those of IL-2R β -deficient mice²⁴, including the defect in T-cell proliferation in response to IL-2, a higher percentage of CD69⁺ T cells, and age-dependent disappearance of thymocytes²⁴. These data suggest that *FADD* is involved in cytokine signalling. IL-2 is implicated in T-cell apoptosis^{25–28} as well as being involved in proliferation. Production of IL-2 in response to T-cell receptor stimulation leads to proliferation

as well as sensitization to Fas-mediated apoptosis^{25,26}. Further studies are required to show how *FADD* mediates distinct pathways of death and growth, and the molecular relationship between them. □

Methods

Generation of *FADD*^{-/-} mice and *RAG-1*^{-/-} chimaeras. *FADD*^{+/-} and *FADD*^{-/-} ES cells and mice were generated as described²⁹. Chimaeras were analysed by flow cytometry using monoclonal antibodies (Caltag and Pharmingen).

IL-2 secretion and proliferation assay. Splenocytes or lymph-node cells were stimulated with anti-CD3 ascites (1:3,000 dilution) or PMA (5 ng ml⁻¹) and ionomycin (A23187, 60 ng ml⁻¹) or Con A (5 μ g ml⁻¹). After 14 h incubation at 37 °C, cells were collected, and expression of the α and β subunits of the IL-2 receptor was examined by flow cytometry. To determine IL-2 concentration, the culture supernatants were diluted and assayed by CTLL-2 proliferation bioassay using MTT (3-(4,5-dimethyl-thiazol-2-yl) 2,5-diphenyltetrazolium bromide). For proliferation assay, lymph-node cells or splenocytes were seeded in 96-well culture dishes (2 $\times$ 10⁶ ml⁻¹). Anti-CD3 ascites, PMA plus ionomycin, or Con A was added as described above. After 48 h incubation at 37 °C, [³H]thymidine (1 μ Ci) was then added to each sample. [³H]Thymidine incorporation was quantified 12 h later.

Received 26 November 1997; accepted 5 February 1998.

- Nagata, S. & Golstein, P. The Fas death factor. *Science* **267**, 1449–1455 (1995).
- Matiba, B., Mariani, S. M. & Krammer, P. H. The CD95 system and the death of a lymphocyte. *Semin. Immunol.* **9**, 59–68 (1997).
- Boldin, M. P. *et al.* A novel protein that interacts with the death domain. *J. Biol. Chem.* **270**, 7795–7798 (1995).
- Chinnaiyan, A. M., O'Rourke, K., Tewari, M. & Dixit, V. M. *FADD*, a novel death domain-containing protein, interacts with the death domain of Fas and initiates apoptosis. *Cell* **81**, 505–512 (1995).
- Kischkel, F. C. *et al.* Cytotoxicity-dependent APO-1 (Fas/CD95) associated proteins form a death-inducing signaling complex (DISC) with the receptor. *EMBO J.* **14**, 5579–5588 (1995).
- Zhang, J. & Winoto, A. A mouse Fas-associated protein with homology to the human MORT1/*FADD* protein is essential for Fas-induced apoptosis. *Mol. Cell. Biol.* **16**, 2756–2763 (1996).
- Muzio, M. *et al.* FLICE, a novel *FADD*-homologous ICE/CED-3-like protease, is recruited to the CD95 (Fas/APO-1) death-inducing signaling complex. *Cell* **85**, 817–827 (1996).
- Boldin, M. P., Goncharov, T. M., Goltsev, Y. V. & Wallach, D. Involvement of MACH, a novel MORT1/*FADD*-interacting protease, in Fas/APO-1 and TNF receptor-induced cell death. *Cell* **85**, 803–815 (1996).
- Chinnaiyan, A. M. *et al.* *FADD*/MORT1 is a common mediator of CD95 (Fas/APO-1) and tumor necrosis factor receptor-induced apoptosis. *J. Biol. Chem.* **271**, 4961–4965 (1996).
- Hsu, H., Shu, H.-B., Pan, M.-G. & Goeddel, D. V. TRADD-TRAF2 and TRADD-*FADD* interactions define two distinct TNF receptor 1 signal transduction pathways. *Cell* **84**, 299–308 (1996).
- Bodmer, J.-L. *et al.* TRAMP, a novel apoptosis-mediating receptor with sequence homology to tumor necrosis factor receptor 1 and Fas(Apo-1/CD95). *Immunity* **6**, 79–88 (1997).
- Chinnaiyan, A. M. *et al.* Signal transduction by DR3, a death domain-containing receptor related to TNFR-1 and CD95. *Science* **274**, 990–992 (1996).
- Kitson, J. *et al.* A death-domain-containing receptor that mediates apoptosis. *Nature* **384**, 372–375 (1996).
- Marsters, S. A. *et al.* Apo-3, a new member of the tumor necrosis factor receptor family, contains a death domain and activates apoptosis and NF- κ B. *Curr. Biol.* **6**, 1669–1676 (1996).
- Yang, X., Khosravi-Far, R., Chang, H. Y. & Baltimore, D. Daxx, a novel Fas-binding protein that activates JNK and apoptosis. *Cell* **89**, 1067–1076 (1997).
- Chen, J., Lansford, R., Stewart, V., Young, F. & Alt, F. W. *RAG-2*-deficient blastocyst complementation: An assay of gene function in lymphocyte development. *Proc. Natl Acad. Sci. USA* **90**, 4528–4532 (1993).
- Mortensen, R. M., Conner, D. A., Chao, S., Geisterfer-Lowrance, A. A. & Seidman, J. G. Production of homozygous mutant ES cells with a single targeting construct. *Mol. Cell. Biol.* **12**, 2391–2395 (1992).
- Mombaerts, P. *et al.* *RAG-1*-deficient mice have no mature B and T lymphocytes. *Cell* **68**, 869–877 (1992).
- Chu, K., Niu, X. & Williams, L. T. A Fas-associated protein factor, FAF1, potentiates Fas-mediated apoptosis. *Proc. Natl Acad. Sci. USA* **92**, 11894–11898 (1995).
- Sato, T., Irie, S., Kitada, S. & Reed, J. C. FAP-1: a protein tyrosine phosphatase that associates with Fas. *Science* **268**, 411–415 (1995).
- Stanger, B. Z., Leder, P., Lee, T.-H., Kim, E. & Seed, B. RIP: a novel protein containing a death domain that interacts with Fas/APO-1 (D95) in yeast and causes cell death. *Cell* **81**, 513–523 (1995).
- Ogasawara, J., Suda, T. & Nagata, S. Selective apoptosis of CD4⁺CD8⁺ thymocytes by the anti-Fas antibody. *J. Exp. Med.* **181**, 485–491 (1995).
- Hernandez-Caselles, T. & Stutman, O. Immune functions of tumor necrosis factor. *J. Immunology* **151**, 3999–4012 (1993).
- Suzuki, H. *et al.* Deranged T cell activation and autoimmunity in mice lacking interleukin-2 receptor β . *Science* **268**, 1472–1476 (1995).
- Fournel, S., Genestier, L., Robinet, E., Flacher, M. & Revillard, J.-P. Human T cells require IL-2 but not G_i/S transition to acquire susceptibility to Fas-mediated apoptosis. *J. Immunol.* **157**, 4309–4315 (1996).
- Lenardo, M. J. The molecular regulation of lymphocyte apoptosis. *Semin. Immunol.* **9**, 1–5 (1997).
- Paris, L. V. *et al.* Functional responses and apoptosis of CD25 (IL-2R α)-deficient T cells expressing a transgenic antigen receptor. *J. Immunol.* **158**, 3738–3745 (1997).
- Lenardo, M. J. Interleukin-2 programs mouse alpha beta T lymphocytes for apoptosis. *Nature* **353**, 858–861 (1991).
- Ramirez-Solis, R., Davis, A. C. & Bradley, A. Gene targeting in embryonic stem cells. *Methods Enzymol.* **225**, 855–878 (1993).
- Schmid, L., Uittenbogaart, C. H., Keld, B. & Giorgi, J. V. A rapid method for measuring apoptosis and dual-color immunofluorescence by single laser flow cytometry. *J. Immunol. Methods* **170**, 145–157 (1994).

Acknowledgements. We thank members of the Winoto, Allison, Robey and Raulet groups for discussions; P. Schow for technical assistance; A. Nagy, R. Nagy and W. Abramow-Newerly for the R1 ES cells; R. Jaenisch for the J1 ES cells; and E. Robey, B. Ortiz and H. Kasler for critical reading of the manuscript. This work is supported in part by grants from the NIH, National Science Foundation and the Keck Foundation. J.Z. was supported by an Irvington Institute fellowship and is a Leukemia Society of America special fellow.

Correspondence and requests for materials should be addressed to A.W. (e-mail: winoto@uclink2.berkeley.edu).

Mutations of mitotic checkpoint genes in human cancers

Daniel P. Cahill^{*,†}, Christoph Lengauer^{*,†}, Jian Yu^{*}, Gregory J. Riggins^{*}, James K. V. Willson[‡], Sanford D. Markowitz[‡], Kenneth W. Kinzler^{*} & Bert Vogelstein^{*}

^{*} The Johns Hopkins Oncology Center, Program in Human Genetics, and The Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, 424 North Bond Street, Baltimore, Maryland 21231, USA

[‡] Department of Medicine and Ireland Cancer Center, Case Western Reserve University and The Howard Hughes Medical Institute, Cleveland, Ohio 44106, USA

[†] These authors contributed equally to this study.

Genetic instability was one of the first characteristics to be postulated to underlie neoplasia^{1–3}. Such genetic instability occurs in two different forms. In a small fraction of colorectal and some other cancers, defective repair of mismatched bases results in an increased mutation rate at the nucleotide level and consequent widespread microsatellite instability^{4–7}. In most colorectal cancers, and probably in many other cancer types, a chromosomal instability (CIN) leading to an abnormal chromosome number (aneuploidy) is observed⁸. The physiological and molecular bases of this pervasive abnormality are unknown. Here we show that CIN is consistently associated with the loss of function of a mitotic checkpoint. Moreover, in some cancers displaying CIN the loss of this checkpoint was associated with the mutational inactivation of a human homologue of the yeast *BUB1* gene; *BUB1* controls mitotic checkpoints and chromosome segregation in yeast. The normal mitotic checkpoints of cells

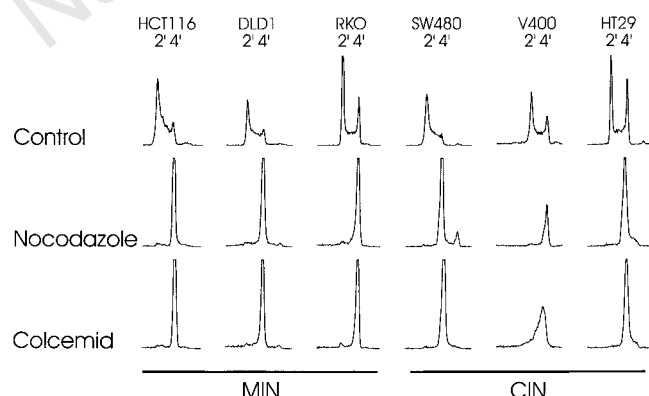


Figure 1 Analysis of the cell cycle of MIN and CIN cells. MIN (HCT116, DLD1, and RKO) and CIN (SW480, V400, and HT29) cells were treated with nocodazole or colcemid for 18 h, stained with Hoechst 33258, a DNA-specific dye, and analysed by flow cytometry. 2' and 4' refer to the DNA contents of each cell line in G1 and G2/M phases, respectively.

displaying microsatellite instability become defective upon transfer of mutant *hBUB1* alleles from either of two CIN cancers.

The key insight leading to the discovery of the molecular basis of microsatellite instability (MIN) in human tumours was the discovery of a similar phenotype in *Saccharomyces cerevisiae* cells carrying mutations in yeast mismatch repair (MMR) genes⁹. Following this paradigm, we reasoned that the basis for CIN in human tumour cells might be mitotic checkpoint defects similar to those previously observed in yeast cells with chromosomal instability^{10–12}. Cells with such defects are expected to exit mitosis prematurely after treatment with microtubule-disrupting agents^{12,13}. To test this hypothesis in human colorectal cancer cells, we treated four MIN lines (HCT116, DLD1, RKO, and SW480) and six CIN lines (SW480, HT29, V400, V429, Caco2, and SW837) with nocodazole, a microtubule-disrupting drug. As expected, all lines achieved nearly complete cell-cycle blocks shortly after nocodazole treatment, with DNA contents of 4C (with 2C representing the DNA content in the G1 phase of the cell cycle, before DNA replication has occurred; representative examples are shown in Fig. 1). Morphological analysis of the 4C blocked cells, however, revealed a striking difference between MIN and CIN cells. All MIN cell lines had a normal checkpoint response, resulting in an accumulation of cells with condensed chromosomes characteristic of a sustained mitotic block. In the CIN lines, there was an abnormal response, with many fewer mitotic cells and no clear peak in mitotic index observed at any time point (Fig. 2a). The response of MIN cells was characteristic of those with intact mitotic checkpoints¹³, and a similar response was observed in normal human fibroblasts (Fig. 2a).

Consistent differences in mitotic indices were also observed in CIN cells versus MIN cells after treatment with colcemid (Figs 1 and 2b), an agent that blocks microtubules through a different mechanism from that of nocodazole. This defect was further established by the higher fraction of CIN cells that synthesized DNA during nocodazole or colcemid treatment (Fig. 2c). There was little overlap between the responses observed in MIN and CIN cells in these

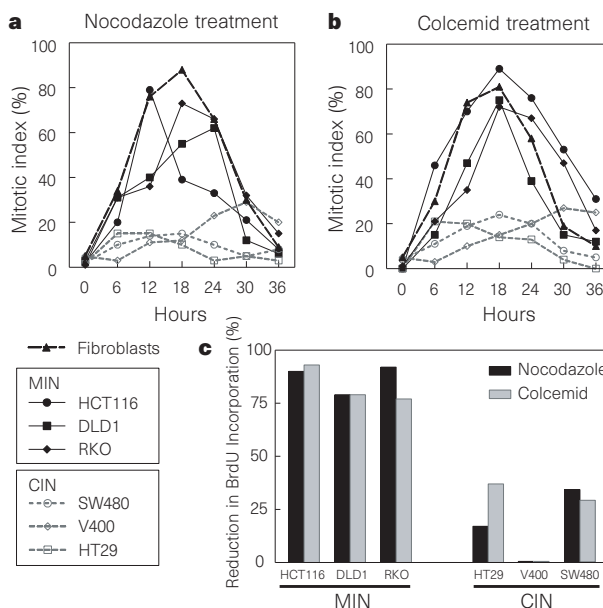


Figure 2 Mitotic indices and DNA synthesis in MIN and CIN cells. Cells were treated with nocodazole (a) or colcemid (b) for the indicated times, stained with H33258, and analysed by fluorescence microscopy. c, BrdU was added to the nocodazole- or colcemid-treated cultures 2.5 h before collection. The bars represent the percentage reduction in BrdU incorporation compared with untreated cells, assessed using an antibody to BrdU. At least 200 cells were counted for each determination and the result shown is representative of those observed in three independent experiments.

assays. For example, the decrease in bromodeoxyuridine (BrdU) incorporation into newly synthesized DNA ranged from 61% to 92% in MIN cells and from 0% to 34% in CIN cells. Similarly, the maximal mitotic indices achieved following 12–18 h of colcemid treatment ranged from 72% to 89% in the four MIN lines, whereas those in the six CIN lines ranged from 20% to 31% ($P < 0.01$ by two-tailed Student's t -test).

To elucidate the genetic mechanisms underlying the checkpoint defect in CIN cells, we chose first to evaluate the human homologue of *S. cerevisiae* *BUB1*. *BUB1* is the prototype member of a family of genes, some of which encode proteins that bind to the kinetochore

and all of which are required for a normal mitotic delay in response to spindle disruption^{14–17}. The human (h) homologue of *BUB1* was cloned and the complete coding sequence determined using a combination of DNA-database searches and reverse transcription with polymerase chain reaction (RT-PCR) methods. Comparison with the *S. cerevisiae* *BUB1* gene showed that these two genes contain two highly conserved domains (CD1 and CD2; Fig. 3a). CD1 (*hBUB1* codons 21–152) directs kinetochore localization and binding to Bub3 whereas CD2 (codons 732–1043) encodes the kinase domain^{15,18}. We isolated a genomic clone containing the *hBUB1* gene and used it to map the location of *hBUB1* to chromosome 2q12–14 through fluorescence *in situ* hybridization (FISH).

We then selected 19 colorectal cancer cell lines with a CIN phenotype to search for mutations in *hBUB1* through RT-PCR-mediated amplification and direct sequencing of the entire coding region of *hBUB1*. RT-PCR analysis of V400 (one of the lines whose phenotype is shown in Figs 1 and 2) revealed a small amplification product (~800 base pairs (bp)) in addition to one of normal size (~1,000 bp). Sequencing of the shorter RT-PCR product showed that it was the result of an internal deletion of 197 bp, predicted to remove codons 76 to 141 and create a frameshift immediately thereafter. Sequencing of the relevant region of genomic DNA identified a G-to-A transition at the canonical splice donor site, which comprises the first intronic nucleotide following codon 140 (Fig. 3b). Sequence analysis of cDNA from another line, V429, revealed a missense mutation at codon 492 (Fig. 3b) that resulted in the substitution of tyrosine for a conserved serine. In both V400 and V429, the mutations were heterozygous and the second allele of the pair was wild type (WT). Analysis of DNA derived from archived tissues of the patients from whom these cell lines were derived revealed that the mutations were somatic, present in their primary tumours but not in their normal tissues (Fig. 3b).

To determine whether the mutant alleles in V400 and V429 could be functionally distinguished from null or WT alleles, we expressed them in MIN cells and evaluated the cells' behaviour following microtubule disruption. Expression vectors encoding both an *hBUB1* cDNA (either WT or one of the two mutants) and a green fluorescent protein (GFP) gene were constructed and transfected into MIN cell lines. We treated cells with nocodazole and productively transfected cells were isolated by cell sorting on the basis of GFP fluorescence (Fig. 4). Over 90% of transfected cells were found to accumulate a DNA content of 4C following nocodazole treatment, regardless of which *hBUB1* sequences were present in the

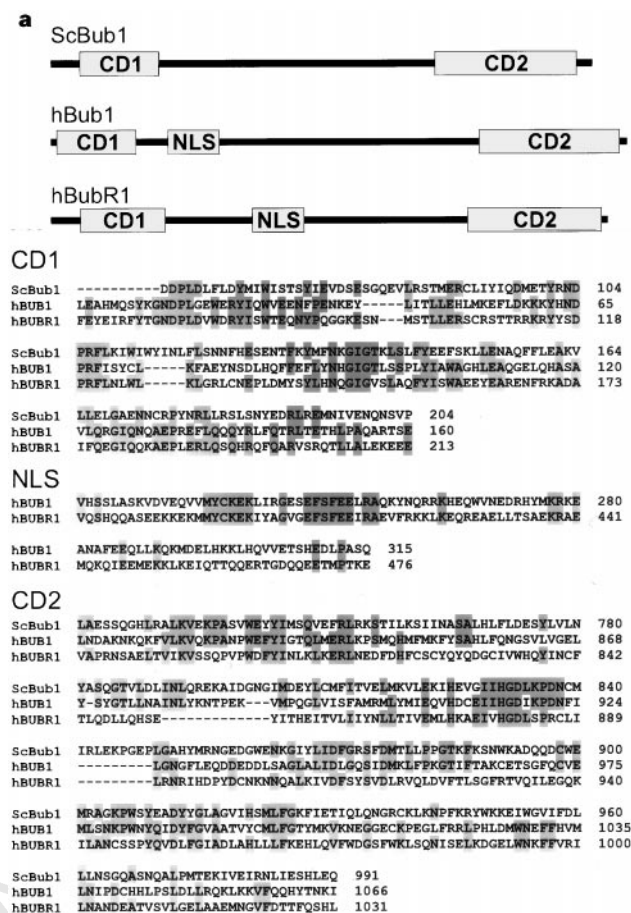


Figure 3 Sequence analysis of human *BUB1* genes. **a**, *S. cerevisiae* (Sc) and human protein sequences were aligned using Macaw Version 2.0.3. CD1, NLS (nuclear-localization signal), and CD2 represent sequence blocks that are highly related among the genes. Identical amino acids are shaded. **b**, Sequence of *hBUB1* in CIN cell lines V400 and V429, in normal cells and in the primary tumours from the patients from whom the corresponding cell lines were derived. Arrows indicate the G-to-A transition at the canonical splice donor site following codon 140 in line V400 and the C-to-A transversion at codon 492 in line V429.

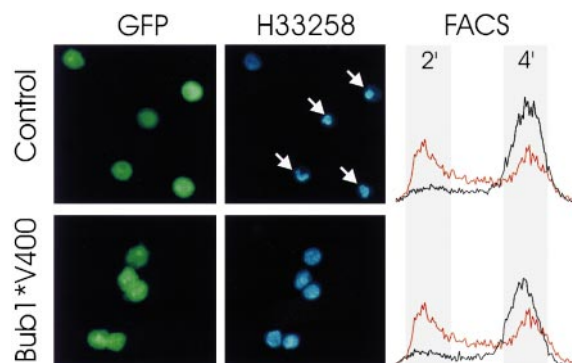


Figure 4 The *hBUB1* expression system. Following transfection of an expression vector encoding mutant *hBUB1* from V400 (Bub1*V400) or of a control vector without *hBUB1*, into HCT116 cells, fluorescence-activated cell sorting (FACS) was used to isolate transfected cells on the basis of GFP expression. Cells were analysed with fluorescence microscopy to ensure that the sorted cells expressed GFP (left panels), and to analyse both mitotic indices following H33258 staining (middle panels) and cell-cycle distributions (right panels). Arrows point to cells in mitosis. The cell-cycle distributions before (red profiles) and after (black profiles) nocodazole treatment are shown on the right. 2' and 4' are defined in Fig. 1.

vector (examples in Fig. 4). However, expression of either of the mutant *hBUB1* genes substantially altered the mitotic indices of transfected cells after nocodazole treatment, whereas the WT *hBUB1* gene had little effect, compared with a control vector lacking *hBUB1* sequences (Fig. 5a). These results indicated that both mutants could confer a dominant-negative effect in two different MIN cell lines, at least when overexpressed relative to endogenous levels. To determine whether a dominant-negative effect was observed when mutant and WT genes were expressed at equivalent levels, we performed cotransfections. Both *hBUB1* mutant expression vectors resulted in altered checkpoint status when transfected with an equal amount of the WT *hBUB1* expression vector; the mitotic index was decreased by 45–69% compared with the index resulting from transfection of the WT *hBUB1* expression vector alone.

If the *hBUB1* mutants caused a premature exit from mitosis in the presence of microtubule disruption, one would expect that such cells would re-enter S phase, just as did the CIN cells depicted in Fig. 2c. To address this point, we added BrdU to the culture media of nocodazole-treated, *hBUB1*-transfected HCT116 cells 2.5 h before collecting them. As shown in Fig. 5b, there was significantly less reduction in BrdU incorporation after transfection with either of the two mutant *hBUB1* expression constructs than after transfection with the WT *hBUB1* expression construct or with the control vector. A similar re-entry into S phase was observed in DLD1 cells after expression of mutant *hBUB1* genes (Fig. 5b).

Because many genes control the mitotic checkpoint in yeast, we searched for additional genes that might play a role in this process in humans. Another *BUB1* homologue, named *hBUBR1*, was identified and its sequence determined through a strategy similar to that described above for *hBUB1* (Fig. 3a). The *hBUBR1* gene is of comparable size to *hBUB1* but is less homologous to the murine *BUB1* gene (29% versus 81% identical residues in the conserved domains, respectively). Sequence comparisons also revealed a third domain between CD1 and CD2 that contains a putative nuclear-localization signal that is present in both human homologues and in murine *BUB1* but not in the yeast gene (Fig. 3a). We isolated a bacterial artificial chromosome (BAC) clone containing *hBUBR1* and used it in FISH experiments to show that the gene is located on human chromosome 15q14–21. We found the *hBUBR1* gene, like *hBUB1*, to be expressed in all 19 CIN cancers analysed. Using RT-PCR products as templates for sequencing, we noted several variants

of *hBUBR1* that are likely to be polymorphisms on the basis their frequency. We also found two mutations, neither one of which was found among 40 normal alleles. The first was a germline C-to-T transition at codon 40, resulting in a substitution of methionine for threonine, in line V531. The second mutation, found in line V1394, was a somatic deletion of T at codon 1,023. Although this latter mutation would be predicted to remove part of the kinase domain within CD2 (ref. 19), further studies will be required to determine whether either of these *hBUBR1* mutations functionally alters the gene product.

The results described above indicate that all tested CIN cell lines were defective in a well defined mitotic checkpoint and had a phenotype similar to that seen in yeast cells with genetic alterations of mitotic checkpoint genes such as *BUB1*. Although such results, in addition to the mutational analyses reported above, do not prove that aneuploidy can be due to defects in mitotic checkpoints, the data are consistent with this possibility. This hypothesis is supported by the fact that the expression of either of two naturally occurring *hBUB1* mutants converted the normal checkpoint status of MIN cells to the defective type characteristic of CIN cells. Analogously, it has been shown that expression of *in vitro*-generated variants of murine *BUB1* allowed cells to exit mitosis prematurely¹⁸. Such dominant-negative effects are intriguing in light of the previous demonstration that the CIN phenotype is dominant upon fusion of MIN and CIN cells⁸.

It will be important to determine whether defects in other mitotic checkpoint genes contribute to mitotic checkpoint defects and aneuploidy in other human cancers. Interestingly, the expression of one such gene (*hSMAD2*) was relatively low in breast cancer cells that exhibited an abnormal mitotic checkpoint²⁰. Finally, the purposeful activation of cell-cycle checkpoints by exogenous agents might be used to exploit the differences between normal and tumour cells described here and thus improve the selectivity of chemotherapy^{16,21,22}.

Note added in proof: A partial cDNA for *hBUB1* (AF011387) has been independently identified by F. Pangilinan *et al.*²⁹.

Methods

Cell culture and cell-cycle analysis. The derivation and growth conditions of the colorectal cancer cell lines used in this study have been described^{8,23,24}. The ten cell lines studied for mitotic checkpoint status were analysed in detail for karyotype and MMR, as described⁸. The four MIN lines (HCT116, DLD1, RKO, and SW48) were MMR-deficient and diploid whereas the six CIN lines were MMR-proficient and aneuploid, with substantial variability in chromosome number and modal chromosome numbers of 71, 119, 82, 86, 96, and 40 in lines HT29, SW480, V400, V429, Caco2, and SW837, respectively. Primary fibroblasts from normal human skin were obtained from Clonetics and used at third passage, when their doubling time was ~26 h, similar to that of the cancer cell lines. Nocodazole or colcemid was added to the media to final concentrations of 0.2 $\mu\text{g ml}^{-1}$ or 1 $\mu\text{g ml}^{-1}$, respectively. Cells were harvested at 6-h time intervals thereafter, then fixed with glutaraldehyde and stained with Hoechst 33258 (H33258). Viable cells were defined as those without fragmented nuclei or other signs of apoptosis. Flow cytometry (>10,000 cells per sample) was used to evaluate the cell-cycle profile and fluorescence microscopy was used to assess the mitotic index (percentage of viable cells arrested in mitosis). In some experiments, BrdU (10 μM) was added to the cells 2.5 h before harvest and BrdU incorporation was evaluated using a monoclonal antibody to BrdU (Boehringer) and a rhodamine-coupled goat anti-mouse secondary antibody (Pierce).

Cloning of *hBUB1* and *hBUBR1*. Full-length protein and nucleotide sequences of *S. cerevisiae* and murine *BUB1* genes were used to search the National Center for Biotechnology Information expressed sequence tag database (dbEST). Multiple overlapping EST clones were identified for the carboxy-terminal ends of the proteins. To derive the remainder of the sequences, rapid amplification of cDNA ends (RACE) was performed using colonic adenocarcinoma RACE-ready cDNA (Clontech). RT-PCR amplification and sequencing from multiple cDNA samples derived from human tissues provided corrections to the contig

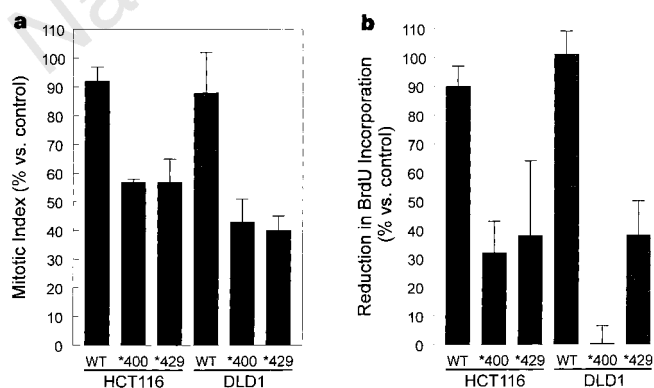


Figure 5 Effects of *hBUB1* expression. Mitotic indices (a) or BrdU incorporation (b) in nocodazole-treated, *hBUB1*-transfected cells were compared with mitotic indices or BrdU incorporation in control-transfected cells. Transfected cells were isolated through cell sorting on the basis of GFP fluorescence. The shaded bars and error bars represent the means and standard deviations, respectively, determined from at least two independent assessments of 100 cells each in a single experiment; similar results were obtained in two independent experiments. The *hBUB1* expression vectors (WT or mutant from V400 [*400] or V429 [*429]), and the cell lines transfected (HCT116 and DLD1), are indicated.

assembled from the databases and EST clones. Genomic clones of each gene were identified by PCR-based screening of a human BAC library (Release III, Research Genetics). FISH was performed on human prometaphase spreads as described²⁵.

Mutational analysis. CIN status of evaluated lines was documented by the presence of aneuploidy or the absence of MIN^{8,26}. Two lines (LoVo and V1394) showed both CIN and MIN, a phenomenon noted previously⁸. Total RNA was purified and used for DNA synthesis with random hexamers and Superscript II (Life Technologies) as described²⁷. The following primers were used to amplify *hBUB1* in three overlapping segments: 5'-CATGGACACCCCGAAAATGTC-3' and 5'-GCATCTTTGCTGGCCACTGC-3' for codons 1–404; 5'-GTGGA-GACATCCCATGAGGATC-3' and 5'-GGATCTTCTGCAATGCGAGCG-3' for codons 304–710; and 5'-AGCCCAAAACAGGCCTTGTCG-3' and 5'-CAGTGTGATTTTAAAGACTGTC-3' for codons 661–1085. The following primers were used to amplify *hBUBR1* in three overlapping segments: 5'-TGATGCTCCGAGGGCAGG-3' and 5'-TTGAGAGCACTCCTACACG-3' for codons 1–249; 5'-GAGTCTTCTGTACCACAACG-3' and 5'-AAAACG-GAGCTGCTTCTGGC-3' for codons 218–617; 5'-TCACAGGCTTCAGAAA-TGTAAC-3' and 5'-ATACATGGTGATCATAAGAGAAC-3' for codons 594–1050. The RT-PCR products were purified on an agarose gel and sequenced with a series of internal primers using Thermosequase (Amersham). The intron–exon structures of selected regions of *hBUB1* and *hBUBR1* were determined through sequence analysis of PCR products generated from genomic DNA. Genomic PCR was performed using primers derived from sequences surrounding the mutations described in the text. DNA from paraffin blocks of the primary tumours from which these lines were derived was purified as described²⁸. The sequences of all primers used for mutational analysis are available from the authors upon request.

***hBUB1* expression vectors.** Expression vector pBI-GFP was constructed by insertion of the blunt-ended *HindIII/NotI* fragment of pHGFP-S65T (Clontech) into the *EcoRV* site of pBI (Clontech). pBI-GFP-Bub1WT was constructed by cloning RT-PCR products representing the WT *hBUB1* sequence into the *NotI* and *SalI* sites of pBI-GFP. pBI-GFP-Bub1*V400 and pBI-GFP-Bub1*V429 were similarly constructed by insertion of RT-PCR products from tumour lines V400 and V429, respectively, into pBI-GFP. HCT116 or DLD1 cells were cotransfected with vector DNA and DNA from a plasmid driving the expression of the tTA transcriptional activator (Clontech) and Lipofectamine Plus (Life Technologies). Twelve hours after transfection, cells were treated with 0.2 µg ml⁻¹ nocodazole. BrdU was added to the media 15.5 h later and the cells were collected 2.5 h after BrdU addition. Successfully transfected cells were isolated by sorting on the basis of GFP fluorescence. Sorted cells were analysed by flow cytometry and fluorescence microscopy as described above. BrdU incorporation was compared with that in transfected cells not treated with nocodazole.

Received 7 January; accepted 6 February 1998.

1. Boveri, T. *Zur Frage der Entstehung maligner Tumoren* vol. 1 (Gustav Fischer, Jena, 1914).
2. Loeb, L. A. Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res.* **51**, 3075–3079 (1991).
3. Hartwell, L. Defects in a cell cycle checkpoint may be responsible for the genomic instability of cancer cells. *Cell* **71**, 543–546 (1992).
4. Perucho, M. Cancer of the microsatellite mutator phenotype. *Biol. Chem.* **377**, 675–684 (1996).
5. Modrich, P. Strand-specific mismatch repair in mammalian cells. *J. Biol. Chem.* **272**, 24727–24730 (1997).
6. Kolodner, R. D. Mismatch repair—mechanisms and relationship to cancer susceptibility. *Trends Biochem. Sci.* **20**, 397–401 (1995).
7. Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colon cancer. *Cell* **87**, 159–170 (1996).
8. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instability in colorectal cancers. *Nature* **386**, 623–627 (1997).
9. Strand, M., Prolla, T. A., Liskay, R. M. & Petes, T. D. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**, 274–276 (1993).
10. Murray, A. W. The genetics of cell cycle checkpoints. *Curr. Opin. Genet. Dev.* **5**, 5–11 (1995).
11. Nasmyth, K. At the heart of the budding yeast cell cycle. *Trends Genet.* **12**, 405–412 (1996).
12. Paulovich, A. G., Toczyski, D. P. & Hartwell, L. H. When checkpoints fail. *Cell* **88**, 315–321 (1997).
13. Nicklas, R. B. How cells get the right chromosomes. *Science* **275**, 632–637 (1997).
14. Li, R. & Murray, A. W. Feedback control of mitosis in budding yeast. *Cell* **66**, 519–531 (1991).
15. Roberts, B. T., Farr, K. A. & Hoyt, M. A. The *Saccharomyces cerevisiae* checkpoint gene BUB1 encodes a novel protein kinase. *Mol. Cell Biol.* **14**, 8282–8291 (1994).
16. Hardwick, K. G., Weiss, E., Luca, F. C., Winey, M. & Murray, A. W. Activation of the budding yeast spindle assembly checkpoint without mitotic spindle disruption. *Science* **273**, 953–956 (1996).
17. Pangilinan, F. & Spencer, F. Abnormal kinetochore structure activates the spindle assembly checkpoint in budding yeast. *Mol. Cell Biol.* **17**, 1195–1208 (1996).
18. Taylor, S. S. & McKeon, F. Kinetochore localization of murine Bub1 is required for normal mitotic timing and checkpoint response to spindle damage. *Cell* **89**, 727–735 (1997).
19. Hanks, S. K. & Quinn, A. M. Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. *Methods Enzymol.* **200**, 38–62 (1991).

20. Li, Y. & Benezra, R. Identification of a human mitotic checkpoint gene: hSMAD2. *Science* **274**, 246–248 (1996).
21. Hartwell, L. H., Szankasi, P., Roberts, C. J., Murray, A. W. & Friend, S. H. Integrating genetic approaches into the discovery of anticancer drugs. *Science* **278**, 1064–1068 (1997).
22. Waldman, T., Lengauer, C., Kinzler, K. W. & Vogelstein, B. Uncoupling of S phase and mitosis induced by anticancer agents in cells lacking p21. *Nature* **381**, 713–716 (1996).
23. Polyak, K., Waldman, T., He, T.-C., Kinzler, K. W. & Vogelstein, B. Genetic determinants of p53 induced apoptosis and growth arrest. *Genes Dev.* **10**, 1945–1952 (1996).
24. Liu, B. *et al.* Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. *Nature Genet.* **9**, 48–55 (1995).
25. Lengauer, C. *et al.* Large-scale isolation of human 1p36-specific P1 clones and their use for fluorescence *in situ* hybridization. *Genet. Anal. Tech. Appl.* **11**, 140–147 (1994).
26. Parsons, R. *et al.* Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer. *Cancer Res.* **55**, 5548–5550 (1995).
27. Riggins, G. J., Kinzler, K. W., Vogelstein, B. & Thiagalingam, S. Frequency of Smad gene mutations in human cancers. *Cancer Res.* **57**, 2578–2580 (1997).
28. Jen, J. *et al.* Allelic loss of chromosome 18q and prognosis in colorectal cancer. *N. Engl. J. Med.* **331**, 213–221 (1994).
29. Pangilinan, F. *et al.* Mammalian BUB1 protein kinases: map positions and *in vivo* expression. *Genomics* **46**, 379–388 (1997).

Acknowledgements. We thank S. Laken for assistance with the microdissection of tumours, J. Flook for assistance with flow cytometry, and T. Henn for assistance with FISH. This work was supported by the Clayton Fund and grants from the NIH. B.V. and S.D.M. are Investigators of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to C.L. (lengauer@welchlink.welch.jhu.edu). The accession numbers for *hBUB1* and *hBUBR1* are AF046078 and AF046079, respectively.

Role of calcineurin and Mpk1 in regulating the onset of mitosis in budding yeast

Masaki Mizunuma, Dai Hirata, Kohji Miyahara, Eiko Tsuchiya & Tokichi Miyakawa

Department of Molecular Biotechnology, Graduate School of Engineering, Hiroshima University, Higashi-Hiroshima 739, Japan

Signalling via calcium is probably involved in regulating eukaryotic cell proliferation, but details of its mechanism of action are unknown^{1,2}. In *Schizosaccharomyces pombe*, the onset of mitosis is determined by activation of a complex of the p34^{cdc2} protein kinase and a cyclin protein that is specific to the G2 phase of the cell cycle. This activation requires dephosphorylation of p34^{cdc2} (ref. 3). Wee1, a tyrosine kinase that inhibits p34^{cdc2} by phosphorylating it, is needed to determine the length of G2 phase. Here we show that calcium-activated pathways in *Saccharomyces cerevisiae* control the onset of mitosis by regulating Swe1, a Wee1 homologue. Zds1 (also known as Oss1 and Hst1) (refs 4–7) is important in repressing the transcription of *SWE1* in G2 phase. In the presence of high calcium levels, cells lacking Zds1 are delayed in entering mitosis. Calcineurin^{8–11} and Mpk1 (refs 12, 13) regulate Swe1 activation at the transcriptional and post-translational levels, respectively, and both are required for the calcium-induced delay in G2 phase. These cellular pathways also induce a G2-phase delay in response to hypotonic shock.

A null mutation of the *ZDS1* gene ($\Delta zds1$) in the W303 *S. cerevisiae* strain causes no obvious phenotype⁴, although a similar mutation ($\Delta oss1 = \Delta zds1$) in other strain backgrounds causes a severe delay in passage through G2 and the cells form a highly elongated bud⁵. However, the $\Delta zds1$ W303 strain did show a growth defect in rich medium (see Methods) supplemented with CaCl₂ (50–300 mM) and the cells formed an elongated bud⁵ (data not shown). These Ca²⁺-induced cell abnormalities were apparently similar to those observed with the $\Delta oss1$ mutant in rich medium without added CaCl₂ (ref. 5). $\Delta zds1$ cells grown in calcium medium were analysed by flow-cytometry analysis (FACS) and staining with 4,6-diamidino-2-phenylindole (DAPI). Most of the elongated cells had one nucleus at the mother/bud neck, which is characteristic of G2 delay (data not shown). FACS analysis showed that the cells had two copies of the DNA, showing that Ca²⁺ causes severe cell-cycle delay in G2 rather than in earlier stages (Fig. 1).

The defect of the $\Delta oss1$ strain can be suppressed by deletion of the *SWE1* gene⁵, which encodes a homologue of *S. pombe* Wee1¹⁴. The Swe1 protein negatively regulates the complex of p34^{CDC28} and Clb2 (a G2 cyclin) by phosphorylating residue Y19 of p34^{CDC28} (ref. 14). We next examined whether deletion of *SWE1* can suppress the Ca²⁺-induced G2 delay in $\Delta zds1$. The $\Delta zds1 \Delta swe1$ cells grew with normal morphology without a G2 delay in calcium medium (Fig. 1), indicating that Swe1 mediates the cell-cycle delay and that Ca²⁺-activated pathway(s) promote cell-cycle delay by positively regulating Swe1.

To examine whether calcineurin is involved in Ca²⁺-mediated cell-cycle delay, we tested the effect of FK506, a potent calcineurin inhibitor^{15,16}. The G2 delay was eliminated by FK506 (data not shown). The Ca²⁺-induced G2 delay was suppressed in the $\Delta zds1 \Delta cnb1$ strain, which lacks functional calcineurin (Fig. 1). We next examined whether Mpk1, a component of a mitogen-activated protein kinase (MAPK) pathway, is also involved in this process, as Mpk1 and calcineurin perform the same functions in some aspects of cell growth¹⁷. Deletion of *MPK1* or *BCK1* (the MAPK kinase of the Mpk1 MAPK pathway) in $\Delta zds1$ cells suppressed the Ca²⁺-induced G2 delay (Fig. 1 and data not shown). Thus, both calcineurin and the Mpk1 pathway probably play essential roles in inducing the G2 delay.

To assess whether these pathways are part of a common mechanism that leads to G2 delay, we examined the effect of overexpressing

MPK1 or *CMP2ΔC*, which encodes a constitutively active form of the calcineurin catalytic subunit. The plasmid containing *MPK1* or *CMP2ΔC*, placed under the control of the galactose-inducible *GAL* promoter, was introduced into the $\Delta zds1$ strain. Cells with pGAL *MPK1* (*GAL::MPK1*) or pGAL *CMP2ΔC* (*GAL::CMP2ΔC*) grew normally in raffinose medium (in which the *GAL* promoter is turned off), whereas in galactose medium (in which the *GAL* promoter is turned on) cell growth was severely inhibited and G2-delayed cells accumulated (Fig. 2a). The two Ca²⁺-triggered pathways probably promote G2 delay by positively regulating Swe1 which then inactivates the p34^{CDC28}/Clb2 complex. To test this we studied the effect of *SWE1* deletion on G2 delay in cells overexpressing *MPK1* or *CMP2ΔC*. As expected, $\Delta zds1 \Delta swe1$ cells bearing pGAL *MPK1* or pGAL *CMP2ΔC* grew normally in galactose medium (Fig. 2b).

How do the Ca²⁺-triggered pathways regulate Swe1? The levels of *SWE1* and *CLN2* messenger RNA oscillate during the cell cycle, and *Oss1/Zds1* may be involved in the periodic repression of transcription of *SWE1* and *CLN2* during the transition from G2 to M phase⁵. To investigate whether the transcription of these genes is regulated by Ca²⁺, we compared *SWE1* mRNA levels during cell-cycle progression in high calcium medium in wild-type and $\Delta zds1$ strains. Cells were arrested and synchronized in G1 phase with α -factor, and levels of *SWE1* mRNA after release from arrest were analysed by northern blotting (Fig. 3). In wild-type cells, *SWE1* mRNA was

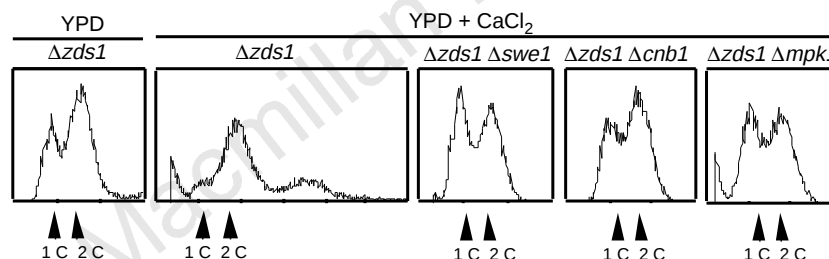


Figure 1 Ca²⁺-induced G2 delay in the $\Delta zds1$ strain and its suppression by deletion of *SWE1*, *CNB1* and *MPK1*. The cells were grown in YPD or YPD

containing 300 mM CaCl₂ at 28°C for 6 h and the DNA content of PI-stained cells was determined by FACS analysis. 1C, one DNA copy; 2C, two DNA copies.

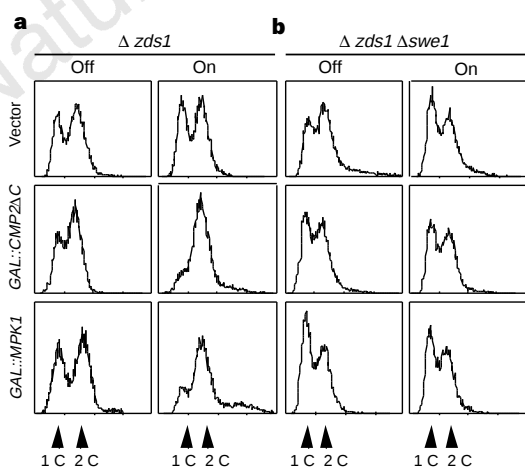


Figure 2 The G2 delay induced by overproduction of a constitutively active form of calcineurin (Cmp2ΔC) or Mpk1 in $\Delta zds1$ and $\Delta zds1 \Delta swe1$ cells. **(a)** $\Delta zds1$ and **(b)** $\Delta zds1 \Delta swe1$ cells carrying pGAL-*CMP2ΔC* (pYES2::*CMP2ΔC*), pGAL-*MPK1* (pNV7::*MPK1*) or vector alone (pYES2) in synthetic complete medium (SD-Ura) were transferred to YPR (raffinose medium, *GAL* promoter turned off) or YPG (galactose medium, *GAL* promoter turned on) and the cells were incubated at 28°C for 10 h. The DNA content (1C or 2C) of PI-stained cells was determined by FACS analysis.

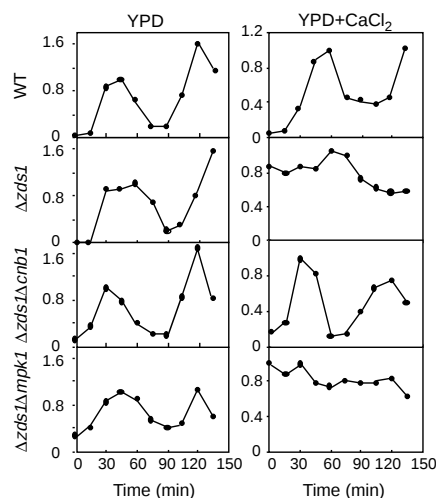


Figure 3 Ca²⁺-induced activation of transcription of *SWE1* in G2. Oscillation of *SWE1* mRNA levels during cell-cycle progression in various strains that were synchronized with α -factor was determined by northern blotting using *SWE1* and *ACT1* probes. In each case, the individual first-cell-cycle maximum value is referred to as 1. WT, wild type. 0 min represents the point of release from G1-phase arrest.

induced 15 min after release from arrest, showing a peak at the G1/S boundary. The induced period continued for longer in Ca^{2+} medium (from 15 to 90 min after release) than in rich medium (from 15 to 75 min after release). In rich medium, $\Delta zds1$ and wild-type cells showed similar oscillation patterns, except that the induced period continued for longer in $\Delta zds1$ cells (from 15 to 90 min after release). In Ca^{2+} medium, $\Delta zds1$ cells showed an altered oscillation pattern. High levels of *SWE1* mRNA continued throughout the cell cycle with no detectable decreases, indicating that the Ca^{2+} leads to sustained *SWE1* transcription in this cell background. Next, we examined the effect of deletion of *CNB1* or *MPK1* on *SWE1* transcription (Fig. 3). In Ca^{2+} medium, sustained *SWE1* transcription occurred in the $\Delta zds1 \Delta mpk1$ strain as in the $\Delta zds1$ strain. In contrast, the $\Delta zds1 \Delta cnb1$ strain displayed a normal

pattern of levels of the *SWE1* transcript. Thus, calcineurin, not Mpk1, is required for the sustained hyperinduction of transcription of *SWE1*. As levels of the *SWE1* transcript oscillate even in the absence of calcineurin, other, unknown factors must be responsible for the cell-cycle-dependent regulation of *SWE1* mRNA, and the calcineurin-regulated auxiliary mechanism must be important in particular conditions. The transcription factor responsible for the calcineurin-mediated G2-phase cell-cycle control is unknown.

We used a genetic approach to ask whether activation of Swe1 by Mpk1 is post-translational. The *HSL1/NIK1* gene encodes a negative regulator (a homologue of *S. pombe* nim1 kinase) of Swe1 (refs 5, 18). The $\Delta hsl1$ strain showed similar cell abnormalities in Ca^{2+} medium to the $\Delta zds1$ strain (data not shown). If Mpk1 regulates Swe1 through Hsl1, overexpression of *MPK1* in the $\Delta hsl1$ strain should not promote G2 delay. Indeed, no G2 delay was seen under these conditions, but G2 delay was induced by *CMP2ΔC* overexpression even in the absence of Hsl1 (Fig. 4a). Furthermore, whereas the $\Delta hsl1 \Delta mpk1$ and $\Delta hsl1$ strains showed a similar G2 delay in high Ca^{2+} medium, there was no G2 delay in the $\Delta hsl1 \Delta cnb1$ strain (Fig. 4b). These results indicate that Mpk1, but not calcineurin, activates Swe1 post-translationally by downregulation of Hsl1, a negative regulator of Swe1. Whether Mpk1 regulates Hsl1 directly or indirectly is unknown.

Calcineurin responds to extracellular high Ca^{2+} levels *in vivo*¹⁹. To determine whether Mpk1 can be activated by extracellular Ca^{2+} , we measured the activity of Mpk1 in cells grown in high Ca^{2+} medium. We used an immune-complex-kinase assay²⁰ with haemagglutinin-tagged Mpk1 (Mpk1^{HA}) and with myelin basic protein (MBP) as a substrate. Mpk1 in wild-type cells was activated by extracellular Ca^{2+} and similar activation occurred in $\Delta zds1$ cells (data not shown). These results indicate that Zds1 is probably unlikely to be an upstream negative regulator of the Mpk1 pathway in response to Ca^{2+} . A model of the cell-cycle control mediated by the Ca^{2+} -signalling pathways is shown in Fig. 5.

The Ca^{2+} -mediated G2 delay is seen only under restricted genetic conditions, such as in the absence of functional Zds1. Nevertheless, Ca^{2+} induces an increase in transcription of *SWE1* and decelerates cell-cycle progression in G2, even in the *ZDS1* background (Fig. 3 and data not shown), indicating a link between the Ca^{2+} signal and the control of G2 in wild-type cells. What is the physiological role of Ca^{2+} -mediated cell-cycle control and which cellular changes are monitored by the Ca^{2+} signal? Yeast cells respond to hypotonic

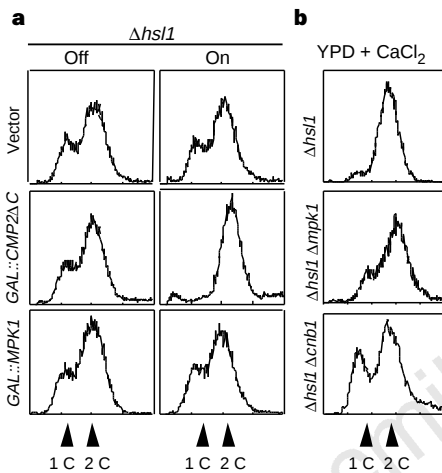


Figure 4 Hsl1-mediated regulation of Swe1 by Mpk1. **a**, the effect of overexpression of *CMP2ΔC* or *MPK1* on G2 delay in the $\Delta hsl1$ strain. The $\Delta hsl1$ strain was transformed with pGAL *CMP2ΔC*, pGAL *MPK1* or vector alone. Experimental conditions were as described in Fig. 2. **b**, the effect of deletion of *MPK1* or *CNB1* on the Ca^{2+} -induced G2 delay of the $\Delta hsl1$ strain. $\Delta hsl1$, $\Delta hsl1 \Delta cnb1$ and $\Delta hsl1 \Delta mpk1$ strains were grown in YPD containing 500 mM CaCl_2 at 28°C for 10 h, and DNA content was analysed by FACS. A higher concentration of CaCl_2 than normal was used, as the $\Delta hsl1 \Delta mpk1$ strain was more resistant to Ca^{2+} than the $\Delta hsl1$ strain in terms of cell abnormality.

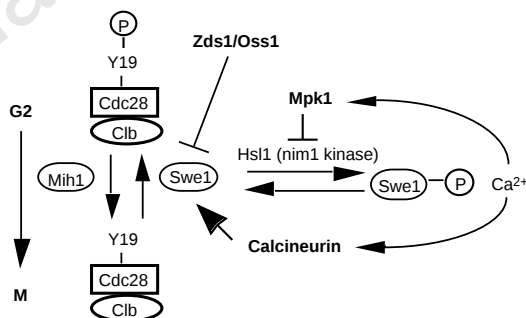


Figure 5 A model of the Ca^{2+} -mediated G2-phase delay in budding yeast. Exit from G2 phase into M phase is controlled by dephosphorylation of the Y19 residue of Cdc28 (which forms a complex with G2 cyclins, or Clbs). Mih1 and Swe1 control the dephosphorylation and phosphorylation, respectively, of Cdc28. The G2 delay is promoted by the positive regulation of Swe1. To increase Swe1 activity, calcineurin induces sustained transcription of *SWE1*, presumably by regulating a transcription factor, whereas the Mpk1 pathway downregulates the activity of Hsl1, which inhibits Swe1 by phosphorylating it. Zds1/Oss1 may inhibit transcription of *SWE1* directly⁵. Alternatively, Zds1 may negatively regulate *SWE1* transcription through inhibition of calcineurin (not shown). Circled P, phosphorylation.

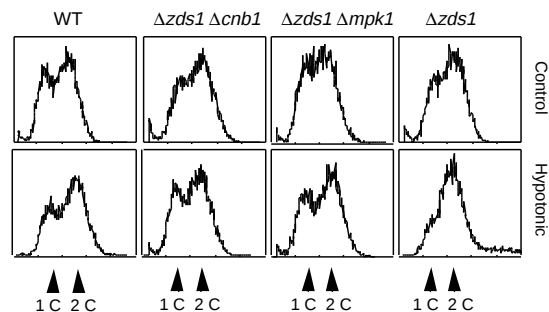


Figure 6 Hypotonic-shock-induced G2 delay. The $\Delta zds1$, $\Delta zds1 \Delta cnb1$, $\Delta zds1 \Delta mpk1$ and wild-type strains were grown in YPD containing 0.8 M sorbitol at 28°C. The cells in the logarithmic phase of growth were exposed to hypotonic shock²² by collecting on a membrane filter and shifting to YPD (hypotonic shock) or YPD containing 0.8 M sorbitol (control). DNA content (1C or 2C of PI-stained cells) was determined by FACS analysis after 1 h of incubation at 28°C. WT, wild-type.

shock with a Ca^{2+} pulse²¹ and transient activation of the Mpk1 pathway, eventually inducing the cellular responses required for growth at the lower osmolarity, such as cell-wall remodelling²². We examined the effect of hypotonic shock on cell-cycle progression (Fig. 6). FACS analysis showed that a strong G2 delay is induced after exposure of $\Delta zds1$ cells, but not $\Delta zds1 \Delta cnb1$ or $\Delta zds1 \Delta mpk1$ cells, to hypotonic shock. A similar, but less prominent, G2 delay was also observed with wild-type (*ZDS1*) cells. Chlorpromazine, which affects membrane biophysics and channel permeability by preferentially inserting into the inner leaflet of the plasma membrane²³, is a potent activator of Mpk1 (ref. 20). Chlorpromazine induced a strong calcineurin- and Mpk1-mediated G2 delay (data not shown). An alteration in membrane biophysics, such as stretching of the plasma membrane, is a likely physiological signal that leads the cells to a delay in G2 phase, through the activation of the Ca^{2+} -signalling pathways, to coordinate cell-cycle progression with cell growth under certain stringent growth conditions. □

Methods

Strains and media. All yeast strains used were derivatives of strain W303 (mating-type *a* *trp1 leu2 ade2 ura3 his3 can1-100*, from R. Rothstein). The yeast strains were YAT1 (*zds1::TRP1*), YMM6 (*zds1::TRP1 swe1::HIS3*), YMM2 (*zds1::TRP1 cnb1::HIS3*), YMM3 (*zds1::TRP1 mpk1::HIS3*), YMM17 (*zds1::TRP1 bck1::HIS3*), YMM12 (*hsl1::URA3*), YMM14 (*MARα hsl1::URA3 cnb1::HIS3*) and YMM15 (*hsl1::URA3 mpk1::HIS3*). Rich medium consisted of 1% yeast extract, 2% polypeptone and 2% dextrose (YPD). Synthetic complete media SD, SG and SR contain 2% each of dextrose, galactose plus raffinose, and raffinose, respectively.

Flow cytometry. Exponentially growing yeast cells ($\sim 1 \times 10^7$) were collected by centrifugation, suspended in 0.2 M Tris-HCl (pH 7.5) and fixed with 70% (v/v) ethanol at 4°C. The fixed cells were washed with the same buffer, sonicated with minimum output for 5 s, suspended in the same buffer containing RNase A (1 mg ml⁻¹) for 4 h at 30°C. The resulting cells were stained with 100 µl of propidium iodide (PI) solution (50 µg ml⁻¹) in 4 mM sodium citrate, 10 mM NaCl and 0.1% NP-40 for 15 min on ice. The cells were analysed by FACScan (ESP flow cytometer, Coulter Epics Elite).

Northern blot analysis. Northern-blot analysis was performed as described^{5,24}. Briefly, cell cultures were synchronized by treating with synthetic α -factor (5 µg ml⁻¹) and washing with water to remove α -factor. The washed cells were suspended in YPD containing 50 mM CaCl_2 . For $\Delta zds1 \Delta mpk1$ cells, 3 µg ml⁻¹ of α -factor peptide was used, as the $\Delta mpk1$ strain is more sensitive to α -factor treatment, in terms of viability, than are other strains. Total RNA was isolated by the hot-phenol method. The isolated RNA was separated on a 1%-agarose gel, transferred to a nylon membrane, and then subjected to northern-blot analysis. The intensity of the bands was measured using a BAS-1000 Bioimaging analyser, and the *SWE1* mRNA levels were normalized to the *ACT1* mRNA levels.

Immune complex protein kinase assay for Mpk1. Preparation of cell extracts, immunoprecipitation, immunodetection and kinase assays were performed as described^{20,25}.

Received 11 November; accepted 12 December 1997.

- Means, A. R. Calcium, calmodulin and cell cycle regulation. *FEBS Lett.* **347**, 1–4 (1994).
- Clapham, D. E. Calcium signaling. *Cell* **80**, 259–268 (1995).
- Nurse, P. Universal control mechanism regulating onset of M-phase. *Nature* **344**, 503–507 (1990).
- Tsuchiya, E. *et al.* The *Saccharomyces cerevisiae* *SSD1* gene is involved in the tolerance to high concentration of Ca^{2+} with the participation of *HST1/NRC1/BFR1*. *Gene* **176**, 35–38 (1996).
- Ma, X.-J., Lu, Q. & Grunstein, M. A search for proteins that interact genetically with histone H3 and H4 amino termini uncovers novel regulators of the *SWE1* kinase in *Saccharomyces cerevisiae*. *Genes Dev.* **10**, 1327–1340 (1996).
- Yu, Y. *et al.* Mutation in the homologous *ZDS1* and *ZDS2* genes affect cell cycle progression. *Mol. Cell Biol.* **16**, 5254–5263 (1996).
- Bi, E. & Pringle, J. R. *ZDS1* and *ZDS2*, genes whose products may regulate Cdc42p in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **16**, 5264–5275 (1996).
- Liu, Y. *et al.* The *Saccharomyces cerevisiae* genes (*CMP1* and *CMP2*) encoding calmodulin-binding proteins homologous to the catalytic subunit of mammalian protein phosphatase 2B. *Mol. Gen. Genet.* **227**, 52–59 (1991).
- Kuno, T. *et al.* cDNA cloning of a calcineurin B homolog in *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.* **180**, 1159–1163 (1991).
- Cyert, M. S., Kunisawa, R., Kaim, D. & Thorner, J. Yeast has homologs (*CNA1* and *CNA2* gene products) of mammalian calcineurin, a calmodulin-regulated phosphoprotein phosphatase. *Proc. Natl Acad. Sci. USA* **88**, 7376–7380 (1991).
- Cyert, M. S. & Thorner, J. Regulatory subunit (*CNB1* gene product) of yeast Ca^{2+} /calmodulin-

- dependent protein phosphatase is required for adaptation to pheromone. *Mol. Cell Biol.* **12**, 3460–3469 (1992).
- Errede, B. & Levin, D. E. A conserved kinase cascade for MAP kinase activation in yeast. *Curr. Opin. Cell Biol.* **5**, 254–260 (1993).
- Levin, D. E. & Errede, B. The proliferation of MAP kinase signalling pathways in yeast. *Curr. Opin. Cell Biol.* **7**, 197–202 (1995).
- Booher, R. N., Deshaies, R. J. & Kirschner, M. W. Properties of *Saccharomyces cerevisiae* *wee1* and its differential regulation of p34^{CDC28} in response to G1 and G2 cyclins. *EMBO J.* **12**, 3417–3426 (1993).
- Foor, F. *et al.* Calcineurin mediates inhibition by FK506 and cyclosporin of recovery from α -factor arrest in yeast. *Nature* **360**, 682–684 (1992).
- Nakamura, T. *et al.* Protein phosphatase type 2B (calcineurin)-mediated, FK506-sensitive regulation of intracellular ions in yeast is an important determinant for adaptation to high salt stress conditions. *EMBO J.* **12**, 4063–4071 (1993).
- Nakamura, T., Ohmoto, T., Hirata, D., Tsuchiya, E. & Miyakawa, T. Genetic evidence for the functional redundancy of the calcineurin- and Mpk1-mediated pathways in the regulation of cellular events important for growth in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **251**, 211–219 (1996).
- Tanaka, S. & Nojima, H. Nkl1: a Nim1-like protein kinase of *S. cerevisiae* interacts with the Cdc28 complex and regulates cell cycle progression. *Genes to Cells* **1**, 905–921 (1996).
- Mazur, C. *et al.* Differential expression and function of two homologous subunits of yeast 1,3- β -D-glucan synthase. *Mol. Cell Biol.* **15**, 5671–5681 (1995).
- Kamada, Y., Jung, U. S., Piotrowski, J. & Levin, D. E. The protein kinase C-activated MAP kinase pathway of *Saccharomyces cerevisiae* mediates a novel aspect of the heat shock response. *Genes Dev.* **9**, 1559–1571 (1995).
- Batiza, A. F., Schulz, T. & Masson, P. H. Yeast respond to hypotonic shock with a calcium pulse. *J. Biol. Chem.* **271**, 23357–23362 (1996).
- Davenport, K. R., Sohaskey, M., Kamada, Y., Levin, D. E. & Gustin, M. C. A second osmosensing signal transduction pathway in yeast. *J. Biol. Chem.* **270**, 30157–30161 (1995).
- Martinac, B., Adler, J. & Kung, C. Mechanosensitive ion channels of *E. coli* activated by amphipaths. *Nature* **348**, 261–263 (1990).
- Igual, J. C., Johnson, A. L. & Johnston, L. H. Coordinated regulation of gene expression by the cell cycle transcription factor *SWI4* and the protein kinase C MAP kinase pathway for yeast cell integrity. *EMBO J.* **15**, 5001–5013 (1996).
- Zarrov, P., Mazzoni, C. & Mann, C. The *SLT2 (MPK1)* MAP kinase is activated during periods of polarized cell growth in yeast. *EMBO J.* **15**, 81–91 (1996).

Acknowledgements. We thank M. Grunstein for the *HSL1* disruption plasmid, D. E. Levin for the plasmid containing the gene for HA-tagged Mpk1, and K. Matsumoto and Y. Ohya for discussions.

Correspondence and requests for materials should be addressed to T.M. (e-mail: tmiyaka@ipc.hiroshima-u.ac.jp).

Human β -tryptase is a ring-like tetramer with active sites facing a central pore

Pedro José Barbosa Pereira^{††}, Andreas Bergner^{††}, Sandra Macedo-Ribeiro^{*}, Robert Huber^{*}, Gabriele Matschiner[‡], Hans Fritz[‡], Christian P. Sommerhoff[‡] & Wolfram Bode^{*}

^{*} Abteilung für Strukturforschung, Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, Germany
[‡] Abteilung für Klinische Chemie und Klinische Biochemie in der Chirurgischen Klinik und Poliklinik, Klinikum Innenstadt der Ludwig-Maximilians-Universität, Nußbaumstrasse 20, D-80336 München, Germany

[†] These authors contributed equally to this work.

Human tryptase, a mast-cell-specific serine proteinase that may be involved in causing asthma and other allergic and inflammatory disorders^{1–3}, is unique in two respects: it is enzymatically active only as a heparin-stabilized tetramer, and it is resistant to all known endogenous proteinase inhibitors. The 3-Å crystal structure of human β -tryptase in a complex with 4-amidinophenyl pyruvic acid shows four quasi-equivalent monomers arranged in a square flat ring of pseudo 222 symmetry. Each monomer contacts its neighbours at two different interfaces through six loop segments. These loops are located around the active site of β -tryptase and differ considerably in length and conformation from loops of other trypsin-like proteinases. The four active centres of the tetramer are directed towards an oval central pore, restricting access for macromolecular substrates and enzyme inhibitors. Heparin chains might stabilize the complex by binding to an elongated patch of positively charged residues spanning two adjacent monomers. The nature of this unique tetrameric architecture explains many of tryptase's biochemical properties and provides a basis for the rational design of monofunctional and bifunctional tryptase inhibitors.

Human tryptase (EC 3.4.21.59) is the predominant protein in most human mast cells. It comprises four closely related enzymes called α -, I-, II/ β - and III-tryptase, which share 90–98% sequence identity. With the apparent exception of α -tryptase^{4,5}, these enzymes are activated intracellularly and stored in a catalytically active form in secretory granules. Despite significant sequence similarity with other trypsin-like serine proteinases, tryptase possesses unique properties that are poorly understood (reviewed in refs 1, 6). Tryptase is enzymatically active only in the form of a non-covalent tetramer, which needs to bind heparin or other proteoglycans for stability. In the absence of such acidic polysaccharides (or of high salt concentrations, *in vitro*), tryptase dissociates into inactive monomers rapidly and irreversibly. This process probably restricts its activity *in vivo*^{7–9}. Human tryptase is not inhibited by any known endogenous proteinase inhibitor; its only naturally occurring

inhibitor is the 'atypical' Kazal-type inhibitor leech-derived tryptase inhibitor (LDTI)^{10,11}.

Like trypsin, tryptase preferentially cleaves peptide substrates carboxy-terminal to arginine and lysine residues¹². It efficiently hydrolyses a number of peptide substrates *in vitro* (for example, it hydrolyses the bronchodilatory neuropeptides vasoactive intestinal peptide and peptide histidine methionine). Unlike trypsin, however, tryptase cleaves only a few proteins, inactivating (for example) fibrinogen, fibronectin and high-molecular-weight kininogen, and activating the zymogens of stromelysin-1 and urokinase-type plasminogen activator. With other preformed mediators (such as histamine and proteoglycans), tryptase is released from mast cells in various allergic and certain inflammatory disorders (for example, psoriasis, rheumatoid arthritis, interstitial cystitis, and multiple sclerosis). Tryptase seems to be directly involved in the pathogenesis of asthma, a hypothesis supported by preliminary clinical trials with the synthetic tryptase inhibitor APC366^{3,13}. Attempts to model the structures of tryptases^{2,14,15} have necessarily used monomeric serine proteinases, and cannot predict the tetrameric architecture. To define the tryptase tetramer and obtain a reliable model for rational drug design, we have analysed the X-ray crystal structure of human mast-cell β -tryptase.

Flat, almost square tetramers (Fig. 1) stack in the crystals along the crystallographic 4₁ screw axis. Each edge of the tetramer is in close contact with symmetry-related neighbours, forming infinitely extended sheets in the crystal. Tryptase monomers are positioned at each corner of the tetramer, delimiting a central oval pore that measures roughly 50 Å × 30 Å. The active centres are directed to this pore. Both entrances to this pore are partially occluded by the 147 loop (see below) that projects from each of the monomers (Fig. 1)

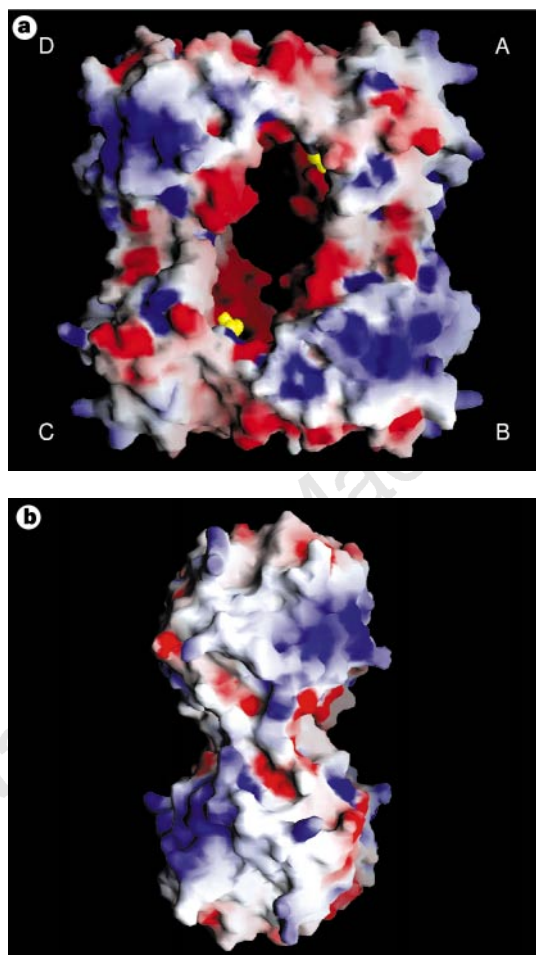


Figure 1 Solid-surface representation of the tryptase tetramer. The colours indicate positive (blue) and negative (red) electrostatic potential at the molecular surface. Figures made with GRASP²⁸. **a**, Front view. The four monomers (labelled A to D) are arranged at the corners of a flat square. Horizontal and vertical local two-fold symmetry axes, relating monomer A (or D) with B (or C), and monomer A (or B) with D (or C), respectively, run through the tetramer centre, as does a third, perpendicular, axis. The four monomers leave a central pore. Two of the four bound APPA molecules (yellow) are visible, whereas the two others are shielded by projecting flaps. The (blue) patches of positively charged residues at the surfaces of monomers B and D extend towards the A-B and C-D peripheries. **b**, Edge view towards the C-D dimer. The tetramer has been rotated around a vertical axis by almost 90° compared with **a**, clearing the view towards the peripheral sides of monomers C (bottom) and D (top). The positively charged patches on both monomers extend towards the front side (right) of D and the back side (left) of C. An extended heparin chain of almost 100 Å could span both patches, strengthening the small C-D and A-B contacts.

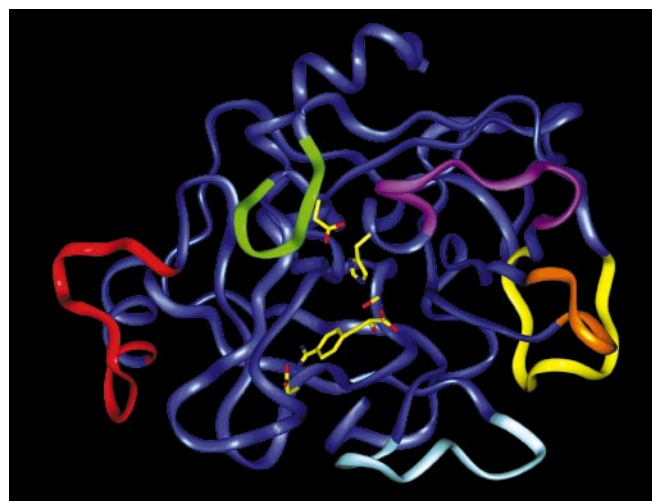


Figure 2 Ribbon representation of one tryptase monomer in the standard orientation. Monomer A of Fig. 1a is shown after rotation around a vertical axis by ~90° (the rotation used to create Fig. 1b). The view is towards the active centre (represented by Ser 195, His 57 and Asp 102 as yellow stick models), with the active-site cleft running from left to right; the APPA molecule interacts with Asp 189 (yellow stick models) in the S1 pocket. The six unique surface loops of tryptase surrounding the active site and engaged in intermonomer contacts are specifically coloured. The 147 loop is light blue; the 70–80 loop is yellow; the 37 loop is orange; the 60 loop is mauve; the 97 loop is green; and the 173 flap is red. All other tryptase segments are violet. Figure produced with Insight II (Biosym/MSI, San Diego).

and restricts access to the wider central cavity. Although it shows (quasi) 222 symmetry, the tryptase tetramer resembles a flat square rather than the compact tetrahedral body shown in most schematic models of tryptase. To our knowledge, a similar flat square tetramer association for a serine proteinase has only been seen in crystals of PR3, an elastase-like enzyme that is stored in, and released from, neutrophil granules¹⁶. In this crystalline 'tetramer', the contacts are mediated through loops that are similar to those in tryptase but differ in the details: the active centres do not face the central pore but are placed opposite to each other in a pairwise manner. However, there is no evidence that PR3 exists as a tetramer in solution.

Not only do the four tryptase monomers, arbitrarily termed A, B, C and D (Fig. 1), 'see' slightly differing crystal environments, but they are also only quasi-equivalent within each tetramer. Monomer A (Fig. 1a) contacts its neighbours B and D through two different interaction surfaces of sizes $\sim 500 \text{ \AA}^2$ and $\sim 1100 \text{ \AA}^2$, respectively; the extent of these contacts (compared with that of the crystal contacts) shows that this complex is the enzymatic and biologically active tryptase tetramer. Monomers A and D (and, likewise, monomers B and C), which are related by an almost exact local twofold axis, are connected by a long peripheral bridge, with both hydrophobic and polar interactions contributing to the intermolecular contacts (Fig. 4b). In contrast, the local twofold symmetry relating monomers A and B (and, likewise, relating C and D) is broken because of the non-equivalent conformation of Tyr 75 (Fig. 4a). Thus, monomer A is equivalent to C, and monomer B is equivalent to D.

The AB homodimer (which is equivalent to CD; Fig. 1b) has several positively charged residues at the periphery; these residues

cluster and form an obliquely oriented two-lobal patch of positive charges that extends towards one of the front sides of each monomer (giving rise to the blue-coloured electrostatic potential surfaces in Fig. 1a, b). This patch, of an overall length of almost 100 Å, would allow tight electrostatic binding of an extended heparin chain of up to 20 sugars, concurring with the observed stabilization of the tetramer by heparin fractions of relative molecular mass 5,500 and above⁷. In contrast, on the peripheral surface of the AD homodimer (and the corresponding BC homodimer), positive charges are counterbalanced by adjacent negative ones.

As in other (chymo)trypsin-like serine proteinases, the core of each tryptase monomer consists of two six-stranded β -barrels, which are attached by three *trans*-domain segments and have two helices and a number of polypeptide loops on their surfaces (Fig. 2). The catalytic residues Ser 195, His 57 and Asp 102 (chymotrypsinogen numbering, Fig. 3) are found at the junction between the two barrels, whereas the active-site cleft runs perpendicular to this junction. In tryptase, 162 or 168 residues are topologically equivalent to the archetypal proteinases chymotrypsin and trypsin, respectively, each with an r.m.s. deviation of their α -carbon atoms of 0.65 Å.

The 15 and 22 residues, which the tryptase exhibits in addition to chymotrypsin and trypsin, respectively, give rise to more exposed

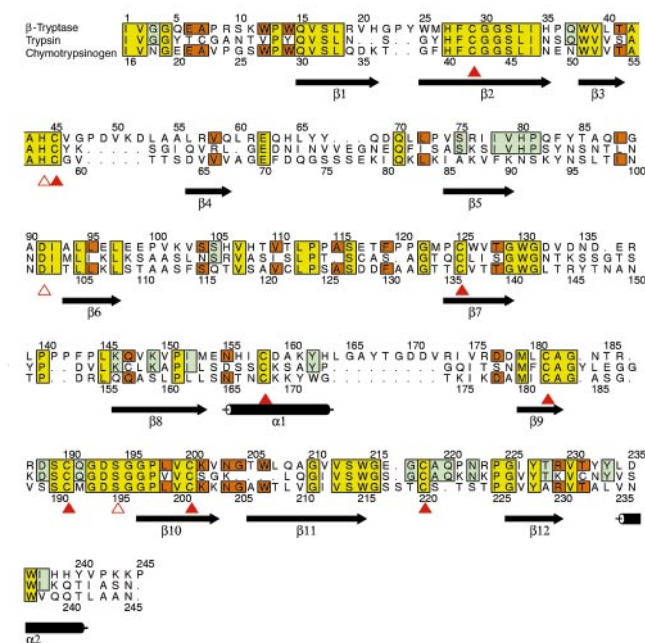


Figure 3 Structure-based amino-acid-sequence alignment. The sequence of human lung mast-cell β -tryptase is aligned with bovine trypsin and the 'B-chain' part of bovine chymotrypsinogen A. Sequence identities between tryptase and trypsin or between tryptase and chymotrypsin are highlighted in green and orange, respectively; identities between all three proteinases are given in yellow. The sequential numbering of tryptase is shown at the top, and the chymotrypsinogen-based numbering used throughout this paper is at the bottom. The catalytic residues and the disulphide-bridge-forming cysteines are marked by open and filled triangles, respectively. The β -strands ($\beta 1$ – $\beta 12$) and α -helices ($\alpha 1$ and $\alpha 2$) of tryptase are indicated by arrows and cylinders, respectively. Figure produced with ALSCRIPT²⁹.

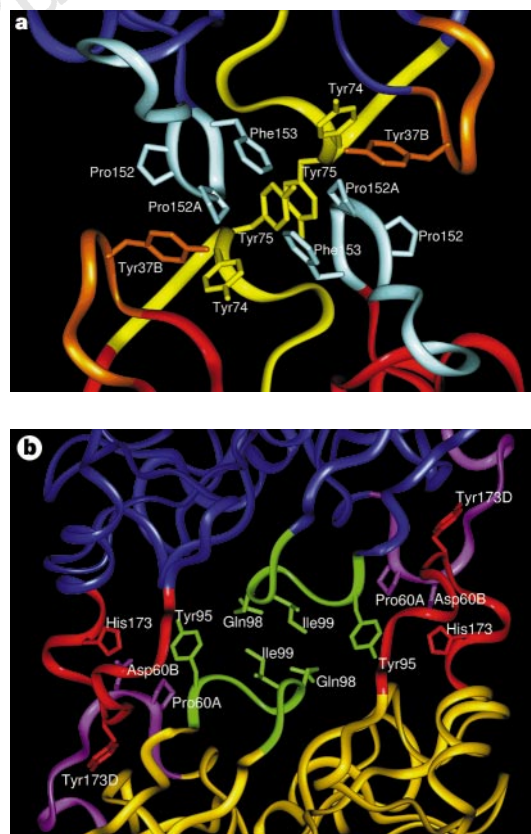


Figure 4 Ribbon representation of the intermonomer contacts. Figures produced with Insight II. **a**, The A–B interface. The view is (as in Fig. 2) along the local A–B two-fold axis, that is, with monomer A (violet) at the top and monomer B (red) at the bottom. The three contacting loops are shown in light blue (152 spur), yellow (70–80 loop) and orange (37 loop). Only a few residues are shown with their side chains. The phenolic groups of Tyr 75 do not obey the twofold symmetry. **b**, A–D interface. The view (from looking down on Fig. 1a) is along the local A–D two-fold axis towards the contact between monomer A (top, violet) and monomer D (bottom, gold). The equivalent 97 loops (green) oppose one another, whereas each 173 flap (red) slots into the groove formed by the opposite 60 loop (mauve). The contact is virtually symmetrical.

and conformationally different loop structures in tryptase than are found in archetypal serine proteinases. This is true particularly for the six surface loops, arranged around the active site, that form the active-site cleft (see Fig. 2 for loop identification and standard orientation). The 60 loop, which contains five inserted residues, turns away from the cleft to the top of the protein where it kinks at *cis*Pro 60A and approaches the general main-chain course of other serine proteinases. Position 69, a strictly conserved glycine residue in most other homologous proteinases, is occupied by a buried arginine residue in tryptase. The next loop, the 70–80 loop, which in the calcium-binding serine proteinases winds around a stabilizing calcium ion¹⁷, is three residues shorter and more compact in tryptase. It is probably not meant to bind calcium, despite topologically similar liganding groups (Glu 70, Asp 80, and carbonyls 72 and 75). The 97 loop, at the top of the cleft, has the same number of residues as in other serine proteinases, but has Ala 97 in the position normally occupied by residue 99, and an unusual helical turn leading to Asp 102. Following the 147 loop, which, with Gln 192, forms the bottom wall of the active-site cleft, the proline-rich sequence Pro 152-Pro 152A-*cis*Pro 152B-Phe 153-Pro 154 forms a hydrophobic 152 'spur'. The largest insertion, of nine residues, occurs in the 173 loop; after the unusually long 3-turn 'intermediate helix' (helix α 1; Figs 2, 3), the ten residues from His 173 to Val 173I form an exposed flap that is centred on the imidazole side chain of His 173 (Fig. 4b).

All monomer–monomer contacts within the tetramer are made by these six loops arranged around each active centre; the formation of these contacts seems to be the main purpose of the unusual loop conformations. Monomers A and B interact with one another through identical loops, namely, the 147 loop, the 70–80 loop, and the 37 loop (Fig. 4a). Each 152 spur slots into a cleft formed by the 37 and the 70–80 loop of its own monomer and the 152 spur of the opposing neighbour. At the centre of the interface, the side chains of Phe 153 and Tyr 75 from each subunit roughly form a tetrahedron. The side chain of Tyr 75 from monomer B (or D)

would clash with the equivalent monomer A (or C) side chain if it were arranged in a symmetrical manner; instead, the phenolic group of Tyr 75 of monomer A turns in the opposite direction, breaking the twofold symmetry (Fig. 4a). This A–B (C–D) interface is exclusively hydrophobic, with several tyrosine and proline side chains being involved.

Monomer A interacts with monomer D through the entire top rim, which consists of the 173 flap, the 97 loop and the 60 loop (Figs 2, 4b), again via equivalent loops. Both 97 loops rest with their 95–99 segments on one another, with both Ile 99 side chains being in direct contact. Towards the periphery, segment Pro 60A–Asp 60B and the opposing segment Gly 173B–Tyr 173D run antiparallel to one another, forming two-rung antiparallel ladders; each aromatic side chain of Tyr 95 nestles into the bend of the opposing 173 flap, and each phenolic side chain of Tyr 173D slots into a hydrophobic cleft made by the 60 loop and the 97 loop of the opposing monomer (Fig. 4b). Both monomers are also cross-connected by salt bridges between Asp 60B and Arg 224 and by four hydrogen bonds. Thus, the A–D (and the corresponding B–C) interface comprises several polar/charged interactions in addition to hydrophobic contacts.

The structure of the vicinity of the tryptase active site is quite similar to that of trypsin. The specificity (S1) pocket, which opens to the left of the reactive Ser 195 residue (Figs 2, 5), is virtually identical to that of trypsin and well suited to accommodating P1 lysine and arginine side chains. The 4-amidinophenylpyruvic acid (APPA) molecule inserts into this pocket in the same way as it inserts into trypsin¹⁸, that is, with its amidino group hydrogen bonded to both Asp 189 carboxylate oxygens, Gly 219 O and Ser 190, and with its phenyl ring sandwiched between peptide planes 215–216 and 190–192 (Figs 2, 5). Ser 195 O γ bonds to the carbonyl group of the tetrahedral pyruvate part of APPA and hydrogen bonds to His 57 Ne.

Below tryptase's active site (Figs 2, 5), the side chains of Asp143 and Asp147 protrude from the relatively flat and hydrophobic cleft embankment. The resulting cluster of negative charges provides a

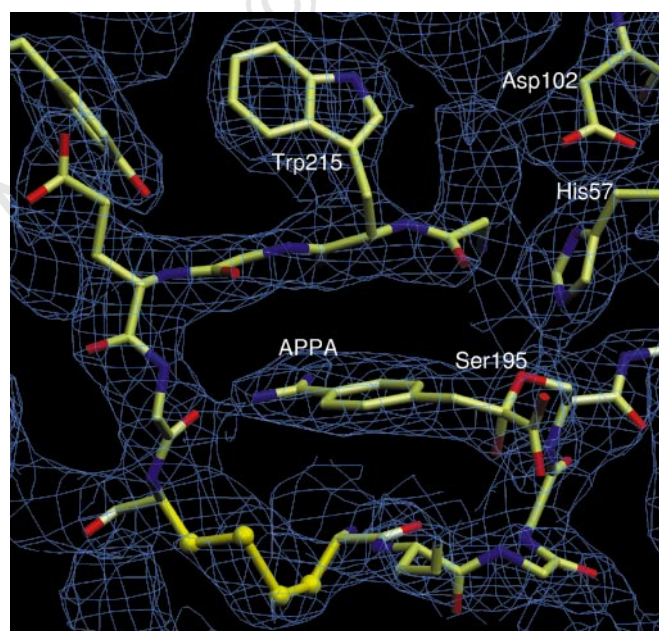


Figure 5 Section of the final 3.0-Å 2Fo-Fc electron-density map (blue) around the specificity pocket of the complex of APPA and tryptase (green). The pocket opens to the left of the active-site triad (right), framed by segment Trp 215-Gly 216-Glu 217-Gly 219-Cys 220, disulphide bridge Cys 220-Cys 191 (yellow), and segment Cys 191-Gln 192-Gly 193-Asp 194-Ser 195. The APPA molecule (centre) is covalently linked to the Ser 195 O γ . Figure produced with SETOR³⁰.

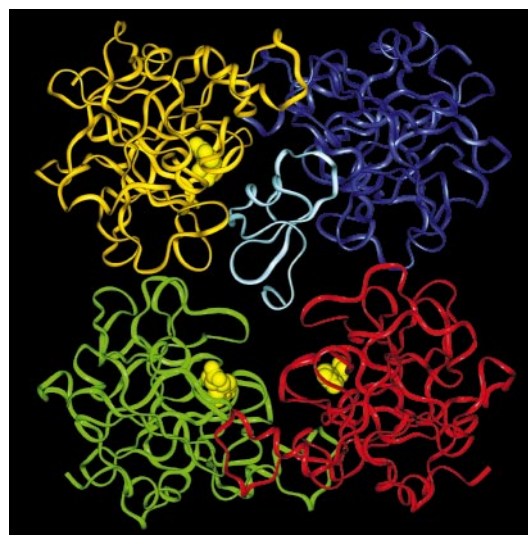


Figure 6 Ribbon diagram of the tryptase-tetramer-LDTI complex. The tetramer, consisting of monomers A (violet), B (red), C (green) and D (gold), is shown in the orientation of Fig. 1a. One LDTI molecule (light blue) is modelled in the active site of monomer A in the way in which it binds to porcine trypsin¹¹, whereas the three other monomers still contain the APPA molecules (yellow). Figure produced with Insight II.

second anchoring point for dibasic synthetic trypsin inhibitors such as bis-benzamidines^{11,19–21}, allowing favourable interactions with a distal basic group. Towards the right, the substrate-binding site of trypsin is bounded not only by the side chains of Tyr 37B and Tyr 74 of monomer A, but also by the benzyl group of Phe 153 and the 152 spur of the neighbouring monomer B, thus limiting the binding of extended substrate chains to about position P5' (Fig. 4a). Near the top, the endogenous 97 loop borders the substrate-binding region in a different way to most other serine proteinases, and, with the side chains of Phe 94, Ala 97 and Gln 98 of monomer D, forms a projecting 'canopy'. The S2 subsite below is open and larger than that of trypsin. The S3/S4 subsite on top of the indole moiety of Trp 215 is fully blocked by the endogenous side chain of Gln 98 and the phenolic group of Tyr 95 provided by monomer D. Towards the left, however, the substrate-binding site is bordered exclusively by segments of the D monomer, in particular by the imidazole ring of His 57 and by segment 57–60. Thus, the active centres of monomers A and D (and those of B and C) are spatially close to each other in the trypsin tetramer, rendering the molecule suitable to specifically bind bifunctional inhibitors with relatively short spacers.

As shown in Fig. 6 for LDTI, the active centre of each trypsin monomer is largely inaccessible to macromolecular substrates and inhibitors. LDTI, an 'atypical' Kazal-type inhibitor that is smaller than the classical members of this family¹⁰, binds to trypsin in a canonical manner through its reactive-site loop (residues P4 to P4')²². In a complex with the A monomer of trypsin, the four amino-terminal residues preceding this binding segment of LDTI could bend towards the bottom (in Figs 2, 5), leading to juxtaposition of the basic Lys1–Lys2 amino terminus of LDTI with the carboxylate groups of Asp 143 and Asp 147. The involvement of such electrostatic interactions is also suggested by the deleterious effect of deletions and substitutions of these basic residues on the inhibitory potency of LDTI¹¹. The pole of LDTI opposite to the reactive-site loop would just touch monomers D and B of β -trypsin both at residue Asp 147. Docking of a second LDTI molecule is only possible at the opposite (C monomer) active centre of the trypsin tetramer; ~50% inhibition of the cleavage activity towards small chromogenic substrates¹⁰ supports this docking theory. Modelling experiments with the more elongated 'classical' Kazal-type inhibitors or with the prototypical bovine pancreatic trypsin inhibitor indicate collisions of the distal poles of these inhibitors with the 147 loop, explaining their inactivity towards trypsin^{1,6}. The larger plasma proteinase inhibitors are too bulky to fit into the narrow trypsin pore and/or its active sites.

The structure of trypsin explains many of its unique properties. It is interesting that access of macromolecular substrates and inhibitors to the active site of trypsin is restricted by the neighbouring monomers rather than by the surface loops surrounding the active site. The unusual conformations of these trypsin loops are required for monomer assembling and tetramer formation. The small and exclusively hydrophobic A–B and C–D contacts seem to be built-in shear points of the tetramer. In the absence of stabilizing heparin chains to join the monomers, or of high salt concentrations to strengthen the hydrophobic A–B contacts *in vitro*, the tetramer disintegrates rapidly, presumably first into (unidentified) dimers A–D and B–C and then into inactive/unfolded monomers, in a multistep process similar to that experimentally observed^{8,9}.

The central pore of trypsin restricts the size of accessible substrates and inhibitors. LDTI, because of its small size, can interact simultaneously with two diagonally located active centres of the trypsin tetramer. Large substrates, such as fibronectin and the zymogens of stromelysin-1 and urokinase-type plasminogen activator, would have to protrude cleavable projections towards the active sites. More flexible neuropeptides, in contrast, could easily thread through this pore to be processed or destroyed. Trypsin is, therefore, well suited for such physiologically important processes as the processing and degradation of neuropeptides. Flexible poly-

peptide chains with two distant basic residues could even dock to adjacent active sites (for example, to active sites of monomers A and D) simultaneously to produce fragments of distinct lengths. The distances between neighbouring active centres are ~20 Å and ~40 Å (A–D and A–B), offering the unique chance of designing bifunctional, trifunctional and tetrafunctional inhibitors to span the gap between the active sites. These multifunctional inhibitors would have higher affinity for trypsin than inhibitors that bind only one active site, thus conferring selectivity for trypsin. Such inhibitors would be useful to pharmacologically probe the pathophysiological function(s) of trypsin *in vivo*, but might also be used to treat asthma³ and other widespread mast-cell-related clinical disorders ranging from allergic conjunctivitis and rhinitis to periodontitis and psoriasis. □

Methods

Protein purification and crystallization. Trypsin was purified from human lung tissue to apparent homogeneity using a modification of described methods⁶. Crystals suitable for diffraction analysis were obtained by sitting-drop vapour-diffusion procedures. The protein in 1.7 M sodium chloride, pH 6.1, was inhibited with excess APPA, and equilibrated at 4 °C against 3 M ammonium sulphate and 0.2 M 3-(N-morpholino)propanesulphonic acid at pH 5.0. Large crystals of the tetragonal space group $P4_1$, with cell axes $a = b = 82.93$ Å, $c = 172.86$ Å, and containing a tetramer per asymmetric unit, grew within 12 months. These crystals initially diffracted to beyond 2.7 Å but showed decay during exposure.

Structure determination. X-ray data to 3-Å resolution were collected from one crystal on a 300 mm MAR Research image plate detector using monochromatized $\text{CuK}\alpha$ radiation from a Rigaku rotating anode X-ray generator. These data were evaluated, merged and scaled with DENZO/SCALEPACK²³, yielding 23,164 independent reflections (99% completeness, $R_{\text{merge}} = 12.1\%$). The structure was solved by molecular replacement with AMoRe²⁴, using a truncated model of porcine pancreatic elastase as the search model. Models were built on an SGI graphics workstation using TurboFRODO²⁵. Calculation of the electron-density maps and crystallographic refinement were performed with REFMAC/CCP4²⁶ and X-PLOR²⁷ using medium and weak non-crystallographic symmetry restraints, respectively, for main- and side-chain atoms.

Present model. In all four monomers, the main chain is fully traceable in electron density from Ile 16 to Pro 243; the following two Lys residues are only partially defined, and the terminal residue, Pro 246, is missing. The model includes 7,716 and 60 non-hydrogen protein and APPA atoms, respectively, and 96 localized solvent molecules. In each monomer, Pro 37A, Pro 60A and Pro 152B are *cis*-prolines. The final R -factor is 19.7% ($R_{\text{free}} = 27.6\%$) including reflections from 10.0 Å to 3.0 Å, and the r.m.s. deviations from target bonds and angles are 0.010 Å and 2.0°, respectively.

Amino-acid sequence. We used the amino-acid sequence of human β -trypsin/trypsin II present in EMBL/PIR databases (entries M37488/B35863 and M33492/P20231, respectively). This sequence was confirmed by digestion of the trypsin used for crystallization with trypsin and mass-spectrometry analysis of the resulting fragments.

Received 28 October; accepted 17 December 1997.

1. Caughey, G. H. *Mast Cell Proteases in Immunology and Biology* (Marcel Dekker, New York, 1995).
2. Caughey, G. H. Of mites and men: trypsin-like proteases in the lungs. *Am. J. Respir. Cell Mol. Biol.* **16**, 621–628 (1997).
3. Seife, C. Blunting nature's swiss army knife. *Science* **277**, 1602–1603 (1997).
4. Schwartz, L. B. *et al.* The alpha form of human trypsin is the predominant type present in blood at baseline in normal subjects and is elevated in those with systemic mastocytosis. *J. Clin. Invest.* **96**, 2702–2710 (1995).
5. Sakai, K., Ren, S. & Schwartz, L. B. A novel heparin-dependent processing pathway for human trypsin. Autocatalysis followed by activation with dipeptidyl peptidase I. *J. Clin. Invest.* **97**, 988–995 (1996).
6. Schwartz, L. B. Trypsin: a mast cell serine protease. *Methods Enzymol.* **244**, 88–100 (1994).
7. Alter, S. C., Metcalfe, D. D., Bradford, T. R. & Schwartz, L. B. Regulation of human mast cell trypsin. Effects of enzyme concentration, ionic strength and the structure and negative charge density of polysaccharides. *Biochem. J.* **248**, 821–827 (1987).
8. Schechter, N. M., Eng, G. Y. & McCaslin, D. R. Human skin trypsin: kinetic characterization of its spontaneous inactivation. *Biochemistry* **32**, 2617–2625 (1993).
9. Addington, A. K. & Johnson, D. A. Inactivation of human lung trypsin: evidence for a re-activatable tetrameric intermediate and active monomers. *Biochemistry* **35**, 13511–13518 (1996).
10. Sommerhoff, C. P. *et al.* A Kazal-type inhibitor of human mast cell trypsin: isolation from the medical

- leech *Hirudo medicinalis*, characterization, and sequence analysis. *Biol. Chem. Hoppe-Seyler* **375**, 685–694 (1994).
11. Stubbs, M. T. *et al.* The three dimensional structure of recombinant leech-derived trypsin inhibitor in complex with trypsin: implications for the structure of human mast cell trypsin and its inhibition. *J. Biol. Chem.* **272**, 19931–19937 (1997).
 12. Kam, C. M. *et al.* Mammalian tissue trypsin-like enzymes: substrate specificity and inhibitory potency of substituted isocoumarin mechanism-based inhibitors, benzamide derivatives, and arginine fluoroalkyl ketone transition-state inhibitors. *Arch. Biochem. Biophys.* **316**, 808–814 (1995).
 13. Clark, J. M., Moore, W. R. & Tanaka, R. D. Trypsin inhibitors: a new class of antiinflammatory drugs. *Drugs Future* **21**, 811–816 (1996).
 14. Johnson, D. A. & Barton, G. J. Mast cell trypsinases: examination of unusual characteristics by multiple sequence alignment and molecular modeling. *Protein Sci.* **1**, 370–377 (1992).
 15. Matsumoto, R., Sali, A., Ghildyal, N., Karplus, M. & Stevens, R. L. Packaging of proteases and proteoglycans in the granules of mast cells and other hematopoietic cells. A cluster of histidines on mouse mast cell protease 7 regulates its binding to heparin serglycin proteoglycans. *J. Biol. Chem.* **270**, 19524–19531 (1995).
 16. Fujinaga, M., Chernaia, M. M., Halenbeck, R., Kothe, K. & James, M. N. The crystal structure of PR3, a neutrophil serine proteinase antigen of Wegener's granulomatosis antibodies. *J. Mol. Biol.* **261**, 267–278 (1996).
 17. Bode, W. & Schwager, P. The refined crystal structure of bovine beta-trypsin at 1.8 Å resolution. II. Crystallographic refinement, calcium binding site, benzamide binding site and active site at pH 7.0. *J. Mol. Biol.* **98**, 693–717 (1975).
 18. Walter, J. & Bode, W. The X-ray crystal structure analysis of the refined complex formed by bovine trypsin and p-aminodiphenylpyruvate at 1.4 Å resolution. *Hoppe-Seyler's Z. Physiol. Chem.* **364**, 949–959 (1983).
 19. Stürzebecher, J., Prasa, D. & Sommerhoff, C. P. Inhibition of human mast cell trypsin by benzamide derivatives. *Biol. Chem. Hoppe-Seyler* **373**, 1025–1030 (1992).
 20. Caughey, G. H., Raymond, W. W., Bacci, E., Lombardy, R. J. & Tidwell, R. R. Bis(5-amidino-2-benzimidazolyl)methane and related amidines are potent, reversible inhibitors of mast cell trypsinases. *J. Pharmacol. Exp. Ther.* **264**, 676–682 (1993).
 21. Fiorucci, L., Erba, F., Bolognesi, M., Coletta, M. & Ascoli, F. pH dependence of bovine mast cell trypsin catalytic activity and of its inhibition by 4',6-diamidino-2-phenylindole. *FEBS Lett.* **408**, 85–88 (1997).
 22. Bode, W. & Huber, R. Natural protein proteinase inhibitors and their interaction with proteinases. *Eur. J. Biochem.* **204**, 433–451 (1992).
 23. Otwinowski, Z. & Minor, W. *DENZO: a Film Processing for Macromolecular Crystallography* (Yale Univ., 1993).
 24. Navaza, J. An automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
 25. Roussel, A. & Cambilleau, C. *TurboFRODO in Silicon Graphics Geometry* (Silicon Graphics, Mountain View, California, 1989).
 26. Collaborative Computational Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
 27. Brünger, G. J. *XPLOR (version 3.1). A System for X-ray Crystallography and NMR* (Yale Univ. Press, 1993).
 28. Nicholls, A., Bharadwaj, R. & Honig, B. GRASP—graphical representation and analysis of surface properties. *Biophys. J.* **64**, A166 (1993).
 29. Barton, G. J. ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37–40 (1993).
 30. Evans, S. V. SETOR: hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graph.* **11**, 134–138 (1990).

Acknowledgements. We thank D. Grosse for help in protein crystallization, R. Mentle for amino-acid-sequence analysis, and M. T. Stubbs for reading the paper. This work was supported by scholarships from Programa Praxis XXI of the Fundação para a Ciência e a Tecnologia and the Programa Gulbenkian de Doutoramento em Biologia e Medicina, Portugal, and by Biotech programs of the European Union, the Sonderforschungsbereich 469 of the University of Munich, the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

Correspondence and requests for materials should be addressed to W.B. (e-mail: bode@biochem.mpg.de). The coordinates have been deposited with the Brookhaven Protein Data Bank under accession number 1A0L.

erratum

Discovery of a supernova explosion at half the age of the Universe

S. Perlmutter^{1,2}, G. Aldering¹, M. Della Valle³, S. Deustua^{1,4}, R. S. Ellis⁵, S. Fabbro^{1,6,7}, A. Fruchter⁸, G. Goldhaber^{1,2}, A. Goobar⁹, D. E. Groom¹, I. M. Hook^{1,10}, A. G. Kim^{1,11}, M. Y. Kim¹, R. A. Knop¹, C. Lidman¹², R. G. McMahon⁵, P. Nugent¹, R. Pain^{1,6}, N. Panagia¹³, C. R. Pennypacker^{1,4}, P. Ruiz-Lapuente¹⁴, B. Schaefer¹⁵ & N. Walton¹⁶

Nature **391**, 51–54 (1998)

The name of one of the authors (Ariel Goobar) of this Letter was inadvertently removed during page make-up. A. Goobar is in the Physics Department, Stockholm University, as indicated in the footnote addresses. □

retracted correction

CTL induction by a tumour-associated antigen octapeptide derived from a murine lung carcinoma

Ofer Mandelboim, Gideon Berke, Mati Fridkin, Michael Feldman, Miriam Eisenstein & Lea Eisenbach

Nature **369**, 67–71 (1994)

A Correction to this Letter was published on 11 December 1997 (*Nature* **390**, 643; 1997), submitted by one of the authors (G.B.). However, because of an error in the editorial office, the correction had bypassed our standard procedures of checking with all authors and, where differences arise, with peer reviewers. The Correction therefore should not have been published. It has since become clear that the matter is one of scientific disagreement between G.B. and the other authors, who stand by their results in the original paper. □

Genetics and genomics

This selection of products includes systems for high-throughput DNA preparation, mutation analysis and fragment screening, a variety of genetics and genomics software, as well as an assortment of probes and labels.

ScanArray 3000

From General Scanning

Microarray biochip technology with laser scanning and micropositioning

This is a fluorescent imaging system designed to capture the full potential of microarray biochips. General Scanning has 30 years of experience in the development of laser systems, and has designed this system to meet the challenges of multi-colour microarray detection and analysis. This system brings easy-to-use Windows software, high-sensitivity, a large dynamic range and a small footprint to microarray detection.

Reader Enquiry No. 100

Gene-Machine

From Rosys Anthos

Fully automates high-throughput DNA preparation

This system combines the Plato robotic microplate processor with the FTA Gene Guard system, and is designed to allow 'walk-away' preparation of purified genomic DNA from biological samples. It will also automate a variety of other protocols, including PCR and post-PCR analysis by ELISA. Building on the efficiency of dried blood stain technology for collecting, transporting and storing, FTA-coated paper has been shown to immobilize DNA within its matrix. Stated to have no carryover, the paper represents a safe and convenient carrier that is well suited for holding DNA for PCR amplification. The Gene-Machine also lends itself to automation and has been developed to perform all stages of the DNA preparation protocol. A fully 'open' system, it offers the flexibility to apply sample to the FTA-coated paper, purify matrix-bound DNA, automatically punch discs of DNA-impregnated paper into a PCR plate, and set up the PCR protocols. With four washable or disposable tips, it handles all sample and reagent distribution and has a robotic arm to transport plates between on-board modules such as a plate washer, drying oven and puncher. Optional barcode readers for both samples and plates offer additional safety and security.

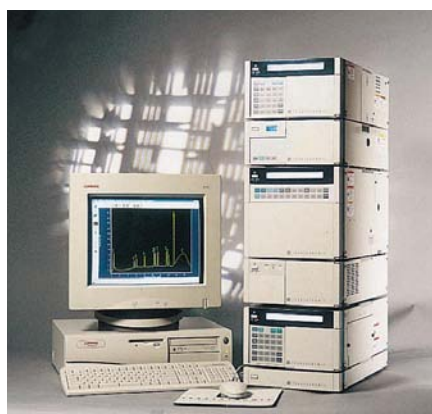
Reader Enquiry No. 101

Wave DNA

From Transgenomic

A system for rapid mutation screening and fragment analysis

This fragment analysis system uses a high-resolution chromatography matrix, DNASep HRM, together with GelPeak software for



Catch the DNA Wave for mutation analysis.

automated sampling and real-time analysis. The system can detect a single-base change in double-stranded DNA of up to 300 bp or single-stranded DNA of up to 60 bp. According to Transgenomic, the purity and quantity of specific DNA fragments can be verified by on-line detection as the fragments are being collected. The technology uses a Peltier-cooled, 96-well autosampler to load samples. The DNA in each sample is quantified and mutations are detected in around five minutes. The GelPeak software shows the results on-screen in a familiar gel format or as chromatographic peaks.

Reader Enquiry No. 102

Software systems

MacVector 6.0

From Oxford Molecular Group

Bioinformatics tools and access to on-line services, offering multiple sequence alignment and improved firewall compatibility

Originally developed by Eastman Kodak and designed specifically to take advantage of the Macintosh computer, MacVector 6.0 is the first release to be developed by Oxford Molecular Group. Of special value to researchers studying sequence data, MacVector incorporates a high-performance implementation of Clustal W and a multi-sequence alignment algorithm for protein and DNA sequences. Results from Clustal W are viewed and edited in the new fully featured alignment editor. The program provides easy access to Entrez sequence databases and supports on-line access to NCBI (National Center for Biotechnology Information), allowing BLAST searches and retrieval of relevant Medline abstracts. Improved firewall compatibility allows on-line access to be exploited by a wider audience. Additionally, MacVector directly provides publication-

quality data output of sequences and alignments without further manipulation for incorporation into reports and scientific papers.

Reader Enquiry No. 103

Cyrillic 2.1

From Cherwell Scientific

This is a fully functional Windows software package for pedigree drawing and genetic data management

Version 2.1 is available as a free demonstration version. The program can be downloaded for testing from the following internet address <www.cherwell.com/cyrillic.html>, or return the journal's reply card to receive a information about a complete demo pack. The demonstration should illustrate how easy it is to create complex pedigrees with a few clicks of the mouse, says Cherwell. It can define multiple chromosomes per family, define multiple diseases and quantitative loci for each family, prepare data for linkage analysis and calculate risks for familial cancers. The manufacturer also offers a 30-day trial period for user testing, during which the product can be returned for a full refund.

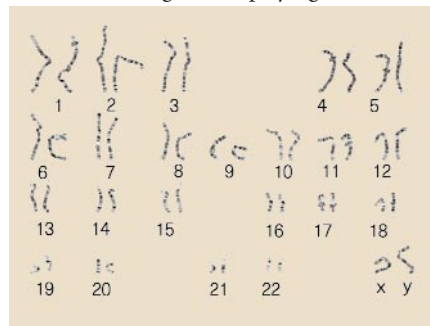
Reader Enquiry No. 104

Quips karyotyping software

From Vysis

Version 3.0 of this integrated clinical, cytogenetic laboratory management system

This karyotyping software includes Quips Lab Manager, an integrated laboratory management system that is also said to provide enhanced accuracy in the automatic classification for G-, R-, Q- and DAPI-banded human chromosomes. It uses computerized imaging technology that automatically captures, segments, classifies and prints high-resolution chromosome images to aid the diagnosis of cytogenetic abnormalities associated with pre- and postnatal disorders and cancer. The upgrade also incorporates new tools for creating and displaying normal and



Quips for human and non-human karyotyping.

abnormal ideograms, improved printing support and additional image enhancement routines. The program automatically classifies and karyotypes non-human species, such as rat and mouse chromosomes. It has more than 60 data fields available for searches, and can operate as a comprehensive cytogenetics laboratory management system. Moreover, it can be interfaced to existing laboratory information systems, if required, and can function as a networked, multi-user system. By using a laptop computer configured with this software, geneticists and physicians can access and view images or patient files from remote locations with appropriate password protection.

Reader Enquiry No. 105

GeneMine

From Molecular Applications Group

An integrated system for bioinformatics that helps to deduce the biological function of a protein from DNA or amino acid sequences

This system is designed to enable the user to go rapidly from discovery of a new gene to the exploration of protein families, prediction of motifs, analysis of structure and identification of function. Hypernotes enable structures and sequences to be electronically linked to text and published as html documents for transmission to collaborators. It filters, clusters and graphically arranges the large volume of information returned from a typical search result, presenting data in a simplified format. According to the manufacturer, this allows users to begin with a comprehensive overview of the data, quickly focus on those relationships among sequences that are likely to yield useful information and then drill down through levels of increasingly fine detail. GeneMine offers data retrieval, integration and visualization capabilities. By accessing available data sources it collects and displays an extensive array of information for the user, including homology clusters, EST tissue expression clusters, homology fingerprints, active sites, functional motifs, post-translational modifications, metabolic pathway associations and sequence polymorphic regions.

Reader Enquiry No. 106

MAR-Finder

From NCGR

A new computer-based gene sequence analysis tool used to identify DNA regions involved in attachment to the nuclear matrix

Available through the centre's Web site at www.ncgr.org/MarFinder/, this tool was developed by senior computational scientist Gautam B. Singh, using statistical inference to deduce the presence of matrix association regions (MARs) in DNA sequences. These constitute a significant functional block within sequences and facilitate differ-

ential gene expression and DNA replication. MAR-Finder is well suited for researchers interested in DNA structure and the identification of genes within a sequence. The tool is said to have been useful in the discovery of the human *SOCS-1* gene. The tool found the matrix association region and then identified the new gene, which is involved in human cytokine-mediated signal transduction. Other new tools available through NCGR include an *ad hoc* query tool, which allows users to query GSDB using structured query language (SQL), and Excerpt, which extracts a selected span or spans from any public sequence in GSDB. An enhanced MAR-Finder is due out in mid-1998.

Reader Enquiry No. 107

GeneWorld 2.5/GeneThesaurus 1.0

From Pangea Systems

Open-system integration for bioinformatics automation software, providing single-point database access and analysis

These systems produce a bioinformatics environment that integrates major sequence and annotation databases through a single point of access. Sold bundled or individually, the open-platform combination enables users to analyse and annotate new sequences automatically, then combine the enriched data with existing databases for cross-referencing and annotation-based analysis. With additional functional enhancements, GeneWorld 2.5 becomes a sequence database application for high-throughput sequence annotation and analysis, enabling users to create automated analysis protocols, called 'Strategies'. Strategies link unknown sequences, algorithms, databases and reports in automated workflow structures that place researchers' analysis throughput capabilities ahead of the daily growth of sequence databases. They can address basic annotations and similarity searching, automated routines for gene finding, functional identification, EST analysis and secondary structure prediction. GeneThesaurus 1.0 links sequence and annotation databases with the information the user has obtained from GeneWorld, and creates a single relational database system. The software cross-references sequence information across databases to uncover significant similarities and creates sets of like sequences based on any annotation field.

Reader Enquiry No. 108

Labelling, probing and expressing

D-FISH DNA probe

From Oncor

For use in minimal residual disease research in chronic myeloid leukemia

This is a new fluorescence *in situ* hybridization (FISH) strategy for detecting the *BCR/ABL1* gene fusion in chronic myeloid



Oncor's D-FISH probe detects *BCR/ABL1* fusion.

leukaemia (CML). Oncor has developed a new probe strategy that helps to eliminate misleading results called D-FISH (dual-fusion FISH). The Oncor *BCR/ABL1* D-FISH translocation DNA probe is formatted to display fusion signals on both the abnormal chromosomes 9 and 22. Combined with strict scoring criteria, this dual fusion strategy reduces the chance of misinterpretation. In minimal residual disease research, this probe can detect low levels of the *BCR/ABL1* gene fusion. The *BCR/ABL1* D-FISH translocation DNA probe is directly labelled in red and green fluorochromes and is available in a ten-test format. The probe is qualified on both peripheral blood and bone marrow.

Reader Enquiry No. 109

RxFISH system

From Applied Imaging

Designed for four-colour chromosome banding analysis based on fluorescent DNA probe technology

Utilizing nucleic acid sequences derived from primate chromosomes, RxFISH allows cancer cytogenetics researcher to characterize better the complex genetic rearrangements that are typically found in certain leukaemia and solid tumours. Current methods of chromosome analysis (including G-banding and colour whole-chromosome painting) are limited in their ability to identify subtle genetic rearrangements, such as inversions that occur within a single chromosome. This method is the first to employ both banding and colour chromosome classification in a single analysis of the whole genome, allowing researchers to identify and characterize genetic rearrangements both within and between chromosomes. This new system comprises a DNA probe reagent kit, a colour chromosome software analysis package and limited hardware upgrades to the company's CytoVision imaging systems.

Reader Enquiry No. 110

Chromoprobe Multiprobe

From Cytocell

A complete range of telomere-specific probes for the detection of all telomeres

This system utilizes a glass device together with a set of fully characterized telomere-specific DNA probes for such uses as the identification of idiopathic or unexplained causes of mental retardation. Some existing

NEW! DNA Ligation, Cloning & Screening in 24 Hours!

Epicentre's new Fast-Link™ DNA Ligation & Screening Kit greatly reduces the time necessary to clone DNA fragments and identify recombinants.

- ❖ **Fast & Easy Ligation.** DNA with restriction enzyme generated cohesive or blunt ends are ligated in just 5 minutes. PCR products with 3' A-overhangs are ligated in 1 hour.
- ❖ **Fast & Easy Screening.** Using Epicentre's exclusive In-Well Gel Screening reagents & protocol, colonies containing recombinant plasmids can be identified quickly without the need to grow cultures or perform minipreps.
- ❖ **High Transformation Efficiencies.** The Fast-Link Kit yields more transformants with a lower background.

Fast-Link™ Representative Results.

	Ligation Time	% White Colonies	Recombinants per µg DNA
Overhang	5 min.	93	2.0 x 10 ⁶
Blunt	5 min.	71	4.4 x 10 ⁵
PCR product	1 hr.	68	1.2 x 10 ⁴

Call for further information on the Fast-Link Kit.

If you are not satisfied with any Epicentre product, we will refund your money.



EPICENTRE™ TECHNOLOGIES
800-284-8474

...when you need to be sure of the quality

Outside of the U.S. contact the distributor in your country, or call 608-258-3080 or fax 608-258-3098.
E-mail: techhelp@epicentre.com
World Wide Web: <http://www.epicentre.com>

methods to detect chromosome abnormalities are inadequate at detecting very subtle changes that occur at the ends of the chromosomes. When DNA probes specific to the telomeres are used in conjunction with the Multiprobe, complicated and subtle changes are said to be detectable. The Chromoprobe Multiprobe T allows the simultaneous analysis of the subtelomeric region of every chromosome for deletion, triplication and balanced translocation events, using a set of 41 telomere-specific DNA probes reversibly bound to an array of raised bosses. Each boss has a probe to the p-arm, which is visualized as a green signal and a probe to the q-arm, which is visualized as a red signal using a fluorescence microscope. The DNA probes used are the first set of fully characterized telomere-specific probes for use in fluorescence *in situ* hybridization (FISH) capable of detecting rearrangements of subtelomeric DNA sequences. Individual telomere-specific probes are also available separately.

Reader Enquiry No. 111

HotPrime

From GenHunter

For radioactive labelling of DNA fragments

According to the manufacturer, this kit yields probes with up to ten times more specific activity than the traditional random priming method. It is designed to maximize the sensitivity of traditional northern or Southern blot analysis by using random decamers instead of hexomers, which are said to be poor primers for DNA polymerase. It also incorporates anchored oligo-dT primers into the labelling buffer to ensure that the 'full-length' antisense cDNA probes are labelled and uses radioactive α -dATP instead of α -dCTP to take advantage of the (A+T)-rich nature of 3' untranslated regions of mRNA. The kit is well suited for labelling probes generated by differential display, but in can also label any other DNA probes to a similar high specific activity.

Reader Enquiry No. 112

BacTen kit

From Quantum Biotechnologies

The kit uses a virus that bears the strong p10 promoter

In BacTen recombinants, the p10 coding sequences and complete polyhedrin locus, including the promoter, is deleted. The p10 promoter remains and is used to express recombinant sequences that are introduced under its control. In this way, expression of both the native polyhedrin and p10 genes is abolished, reducing the background of the very late, native proteins and allowing increased expression of recombinant proteins. Furthermore, the p10 promoter is activated earlier in the course of infection than is the polyhedrin promoter used in

traditional baculovirus vectors, resulting in earlier protein production. Finally, the absence of the p10 protein results in delayed cell lysis, an extended period of recombinant protein synthesis and greater cumulative production of recombinant proteins. This combination of factors is said to increase expression levels (a stated 1.8 times higher in BacTen, than in polyhedrin-controlled expression in wild-type baculovirus (AcMNPV)).

Reader Enquiry No. 113

Database access

Geneseq

From Derwent

A database of nucleic acid and amino acid sequence information from patents, available in flat file (EMBL) format

This should allow research scientists to obtain specific patent data from approximately 270,000 sequence records gleaned from 40 patent issuing authorities. According to the publisher, over half of these sequences are not available on other commonly used sequence databases. It is said to be easy to use, particularly for those not accustomed to information searching. It can be used together with the company's World Patent Index for more comprehensive information on individual patents. The main asset is that the records are prepared by bioinformatics experts and the abstracts focus directly on the novel features and potential application of the sequence itself. Every sequence has an individual record that can be quickly and easily searched to identify particular sequences of interest. Derwent's Geneseq will continue to be available from STN for on-line searching from Oxford Molecular Group for those users who still wish to use the GCG platform in-house.

Reader Enquiry No. 114

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENT

BACs
YACs
PACs
cDNAs

Research Genetics provides clones and colony membranes for a number of human, mouse, and rat libraries as well as a custom screening service for the libraries.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 205-536-9016